PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/92, 15/61

A1

(11) International Publication Number: WO 96/24667

(43) International Publication Date: 15 August 1996 (15.08.96)

(21) International Application Number: PCT/FI96/00070

(22) International Filing Date: 7 February 1996 (07.02.96)

(30) Priority Data:
9502413.9 8 February 1995 (08.02.95) GB

(71) Applicant *(for all designated States except US)*: PRIMALCO LTD [FI/FI]; Valta-akseli, FIN-05200 Rajamäki (FI).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: SUOMALAINEN, Irma [FI/FI]; Maneesikatu 3 C 22, FIN-00170 Helsinki (FI). AHO, Sirpa [FI/US]; 26 Lamp Post Lane, Cherry Hill, NJ 08006 (US). SAARELAINEN, Ritva [FI/US]; 13065 Mindanao Way 17, Marina del Ray, CA 90292 (US). KRISTO, Paulo [FI/FI]; Jyrkännetie 3, FIN-02360 Espoo (FI).

(74) Agent: BORENIUS & CO. OY AB; Arkadiankatu 21 A, FIN-00100 Helsinki (FI).

(81) Designated States: CA, FI, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
With an indication in relation to a deposited microorganism furnished under Rule 13bis separately from the description.
Date of receipt by the International Bureau:
19 February 1996 (19.02.96)

(54) Title: PLANT XYLOSE ISOMERASE

(57) Abstract

The present invention relates to a neucleic acid sequence which codes upon expression in a prokaryotic or eucaryotic host cell for a polypeptide having xylose isomerase activity, which nucleic acid sequence is selected from: a) the nucleic acid sequences shown in SEQ. ID. No. 1 and SEQ. ID. No. 2 or the complementary strands thereof; b) nucleic acid sequences which hybridise to the sequences defined in (a) above; c) nucleic acid sequences which, but for the degeneracy of the genetic code, would hybridise to the sequences defined in (a) or (b) above and which code for the same polypeptides as those defined in (a) or (b) above; and to a process of producing ethanol from lignocellulosic materials comprising contacting cells that express such nucleic acid sequences.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

Plant xylose isomerase

This invention relates to nucleic acid sequences encoding a novel plant xylose isomerase, to vectors carrying the sequences and to the use of the isomerase in a process of ethanol production.

D-xylose is a five-carbon sugar. It is present in nature, for example as xylan polymer in plant hemicellulose. As a carbon source, D-xylose has been reported to be utilized widely by bacteria and to a lesser extent by fungi. D-xylose is first converted to its keto isomer, D-xylulose, which is then phosphorylated to D-xylulose-5-phosphate, a normal pentose phosphate cycle intermediate.

There are two possible routes for the isomerization of D-xylose: (a) direct conversion to D-xylulose by an isomerase; or (b) an oxidoreductive pathway, in which xylitol is an intermediate, and coenzymes for the two oxidoreductases involved are needed. The oxidoreductive pathway has been reported to be the main route in fungi.

Xylose isomerase catalyses the direct isomerization of D-xylose to D-xylulose, and vice versa, with the direction of the reaction dependent on the relative concentrations of the aldo and keto forms and the reaction conditions. Xylose isomerase also catalyses the interconversion of the 5-carbon sugars D-ribose and D-ribulose and the interconversion of the 6-carbon aldose sugar, D-glucose, and its keto isomer, D-fructose. Thus, bacterial xylose isomerases are used industrially to produce D-fructose from D-glucose.

Some eucaryotes have also been found to be capable of converting xylose to xylulose, for example certain yeasts, such as Pichia stipitis. However, the oxidoreductive pathway mentioned above is at least partially responsible for these observations because it has not always been possible to prove

the existence of a xylose isomerase and because oxidoreductive enzymes have been isolated from Pichia stipitis and the other yeasts that have been found to be able to metabolise xylose. In plants, certain tissues are believed to contain xylose isomerase because partially purified extracts of corn pollen and wheat germ are capable of converting xylose to xylulose. This conversion occurs in the absence of NADPH and is not known to require any coenzymes, though Mn^{2+} or Mg^{2+} ions are required.

In order to produce ethanol by fermentation of lignocellulosic material, it is desirable that the organism responsible for the fermentation is able to convert xylose to xylulose. Such organisms generate a high yield of ethanol because they are able to degrade hemicellulosic xylose derived from xylan polymers in lignocellulosic materials. The principal distillers' yeasts, Saccharomyces cerevisiae and Schizosaccharomyces pombe are not naturally able to isomerise D-xylose to D-xylulose although they readily ferment D-xylulose to ethanol.

Some attempts have been made to overcome this problem by transforming S. cerevisiae and S. pombe with bacterial xylose isomerases or oxidoreductive enzymes from yeasts that naturally utilise D-xylose. Also, attempts have been made to enable S. cerevisiae and S. pombe to isomerise D-xylose to D-xylulose by adding bacterial isomerases to fermentation broths.

Yeasts transformed with bacterial isomerases do not produce ethanol efficiently, despite the fact that bacterial xylose isomerases have higher specificity for xylose than for glucose or ribose. It is believed that bacterial isomerase enzymes are poorly expressed in yeast or that they aggregate or are cleaved proteolytically, owing to differences between the procaryotic cells from which they are derived and the eucaryotic yeast host cells.

A eucaryotic xylose isomerase has now been isolated. This is the barley xylose isomerase, which is the first eucaryotic

xylose isomerase to be isolated and therefore the first plant xylose isomerase to be isolated. It is to be expected that the eucaryotic barley enzyme will be more efficiently expressed in yeast than bacterial enzymes because of the genetic similarities between the eucaryotic plant cell from which the enzyme is derived and the eucaryotic yeast cell in which it is expressed.

The amino acid sequence of the enzyme has been determined, the gene encoding it has been cloned and its genomic and cDNA sequences have been determined. Further, it has been found that this barley xylose isomerase has a very high specificity towards D-xylose, as compared to D-glucose and D-ribose. Indeed, in experiments conducted by the present inventors, the barley enzyme was not observed to have any glucose isomerase or ribose isomerase activity. Owing to this high specificity, the enzyme can be used specifically to produce xylulose from xylose, even in the presence of other sugars such as glucose and ribose, as the barley enzyme will not isomerise these other sugars. Accordingly, distillers' yeasts that express the barley xylose isomerase can be used to ensure increased yields of ethanol from xylose in lignocellulosic materials, even in the presence of glucose. Also, the enzyme and cells expressing it, can be used in the modification of xylans.

Accordingly, the present invention provides:

an isolated nucleic acid sequence which codes upon expression in a procaryotic or eucaryotic host cell for a polypeptide having xylose isomerase activity, which nucleic acid sequence is selected from:

a) the nucleic acid sequences shown in SEQ. ID. No. 1 and SEQ. ID. No. 2 or the complementary strands thereof;

b) nucleic acid sequences which hybridise to the sequences defined in (a) above;

c) nucleic acid sequences which, but for the degeneracy of the genetic code, would hybridise to the sequences defined in (a) or (b) above and which code for the same polypeptides as those defined in (a) or (b) above;

a polypeptide encoded by a nucleic acid sequence as defined above;

a vector comprising a nucleic acid sequence as defined above;
cells transformed or transfected with a vector such a vector;
a process of producing ethanol which comprises:

(i) contacting such transformed cells with a substrate that comprises one or more carbon sources selected from xylose and polymerised xylose moieties;

(ii) culturing the said cells in conditions under which the isomerisation of D-xylose to D-xylulose occurs and under which the D-xylulose is further catabolized to ethanol; and

(iii) recovering the ethanol;

an antibody to a polypeptide as defined above;

a process of producing a polypeptide as defined above comprising expressing a nucleic acid sequence as defined above in a cell as defined above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Xylose isomerase activity in barley malted for various lengths of time (Days).

Figure 2. Effect of temperature on enzyme activity.

Figure 3. Effect of pH on enzyme activity.

Figure 4. Exon/intron structure of the barley xylose isomerase gene.

Figure 5. The exon/intron junctions of the barley xylose isomerase gene.

Figure 6. Construction of plasmid pALK710, containing the full lenghth xylose isomerase cDNA.

Figure 7. Construction of plasmids pALK720 and pALK721.

Figure 8. Construction of plasmids pALK724 and pALK725.

Figure 9A. The 5′-terminal end of the cDNA sequence of the full cDNA sequence of barley xylose isomerase disclosed in SEQ. ID. No. 2 and the corresponding part of the full protein sequence of barley xylose isomerase. The total amino acid sequence is disclosed in SEQ. ID. No. 3 and Figures 9A and 9B.

Figure 9B. The 3′-terminal end of the cDNA sequence of the full cDNA sequence of barley xylose isomerase disclosed in SEQ. ID. No. 2 and the corresponding part of the full protein sequence of barley xylose isomerase. The total amino acid sequence is disclosed in SEQ. ID. No. 3 and Figures 9A and 9B.

The nucleic acid sequences of the present invention are preferably DNA, though they may be RNA. It will be obvious to those of skill in the art that, in RNA sequences according to the invention, the U residues shown in SEQ. ID. No. 1-5 will be replaced by T.

The nucleic acid sequences of the present invention are not limited to the sequences of SEQ. ID. No. 1 and SEQ. ID. No. 2. Rather, the sequences of the invention include sequences that are closely related to these sequences and that encode a polypeptide having xylose isomerase activity, xylose isomerase

activity being an ability to catalyse the direct interconversion of xylose and xylulose. These sequences may be prepared by altering those of SEQ. ID. No. 1 or 2 by any conventional method, or isolated from any organism or made synthetically. Such alterations, isolations or syntheses may be performed by any suitable method, for example by the methods of Sambrook et al.: (Molecular Cloning: A Laboratory Manual; 1989).

For example, the sequences of the invention include sequences that are capable of selective hybridisation to those of SEQ. ID. No. 1 and/or SEQ. ID. No. 2 and that encode a polypeptide having xylose isomerase activity. Such sequences capable of selectively hybridizing to the DNA of SEQ. ID. No. 1 and/or 2 will be generally at least 70 %, preferably at least 80 or 90 % and more preferably at least 95 % homologous to the DNA of SEQ. ID. No. 1 or 2 over a region of at least 20, preferably at least 50, for instance 100, 500 or 1000 or more contiguous nucleotides.

Such hybridisation may be carried out under any suitable conditions known in the art (see Sambrook et al. (1989): Molecular Cloning : A Laboratory Manual). For example, if high stringency is required, suitable conditions include 0.2 x SSC at 60 °C. If lower stringency is required, suitable conditions include 2 x SSC at 60 °C.

Also included within the scope of the invention are sequences that differ from those defined above because of the degeneracy of the genetic code and encode the same polypeptide having xylose isomerase activity, namely the polypeptide of SEQ. ID. No. 3, or a polypeptide related to it in any of the ways defined below.

In particular, the non-coding portions of nucleic acid sequences of the invention including the introns of SEQ. ID. No. 1, may be modified in any way that does not destroy the xylose isomerase activity of the encoded polypeptide.

The nucleic sequences of the invention may be of any length as long as they encode a peptide having xylose isomerase activity. For instance, a nucleic acid sequence according to the invention will typically comprise the parts of the native gene sequence that encode the active site of the native protein. A nucleic acid sequence according to the invention may be a contiguous fragment of the native sequence or a sequence that is related to it in any of the ways described above. Alternatively, nucleic acid sequences of the invention may comprise DNA sequences that are not contiguous in the native sequence. These sequences may be fragments of the native DNA sequence or nucleic acid sequences that are related to such fragments in any of the ways described above. Nucleic acid sequences of the invention will preferably comprise at least 50 bases, for example 50 to 100, 100 to 500, 500 to 1000, or 1000 to 2000 bases.

Similarly, the polypeptides of the invention are not limited to the polypeptide of SEQ. ID. No. 3. Rather, the polypeptides of the invention also include polypeptides with sequences closely related to that of SEQ. ID. No. 3 that have xylose isomerase activity. These sequences may be prepared by altering those of SEQ. ID. No. 3 by any conventional method, or isolated from any organism or made synthetically. Such alterations, isolations or syntheses may be performed by any conventional method, for example by the methods of Sambrook et al. (Molecular cloning: A Laboratory Manual; 1989). In particular, polypeptides related to that of SEQ. ID. No. 3 may be prepared by modifying DNA sequences as shown in SEQ. ID. No. 1 or 2 and expressing them recombinantly.

Polypeptides of the invention may include substitutions, deletions, insertions, or extensions that distinguish them from SEQ. ID. NO. 3 as long as these do not destroy the xylose isomerase activity of the polypeptide.

A substitution, deletion or insertion may suitably involve one or more amino acids, typically from one to five, one to ten or

one to twenty amino acids. For example, a substitution, deletion or insertion of one, two, three, four, five, eight, ten, fifteen, or twenty amino acids. Typically, a peptide of the invention has at least 40 %, at least 60 %, at least 80 %, at least 90 %, or at least 95 % sequence identity to native barley xylose isomerase sequence (SEQ. ID. NO. 3.).

In general, the physicochemical nature of the sequence of SEQ. ID. No. 3 should be preserved in a sequence of the invention. Such sequences will generally be similar in charge, hydrophobicity and size to that of SEQ. ID. No. 3. Examples of substitutions that do not greatly affect the physicochemical nature of amino acid sequences are those in which an amino acid from one of the following groups is substituted by a different amino acid from the same group:

H, R and K
I, L, V and M
A, G, S and T
D, E, P and N.

As far as extensions are concerned, a sequence of one or more amino acids may be provided at either or both of the C- and N-termini of the sequence of SEQ. ID. No. 3 or a sequence related to it in any of the ways defined herein. An extension may comprise up to 5, up to 10, up to 20, up to 50, or up to 100 amino acids. For example, an extension may comprise one, two, three, four, five or ten amino acids.

A polypeptide of the invention may be subjected to one or more chemical modifications, such as glycosylation, sulphation, COOH-amidation or acylation.

A polypeptide of the invention may comprise multiple copies of the sequence of SEQ. ID. NO. 3, or a sequence related to it in any of the ways defined herein.

A polypeptide of the invention may be of any length as long as

it has xylose isomerase activity. For instance, a polypeptide of the invention might comprise the active site of the native protein and therefore have xylose isomerase activity despite being much smaller than the native protein. Polypeptides according to the invention may be composed of a contiguous fragment of the native protein sequence or a fragment that is related to it in any of the ways described above. Alternatively, polypeptides of the invention may comprise amino acid sequences that are not contiguous in the native protein. These amino acid sequences may be identical to parts of the native amino acid sequence or related to such parts in any of the ways described above. Polypeptides according to the invention preferably comprise at least 10 amino acids, for example 10 to 20, 20 to 50, 50 to 100, 100 to 200, or 200 to 500 amino acids.

Polypeptides according to the invention may be purified or substantially purified. Such a polypeptide in substantially purified form will generally comprise the polypeptide in a preparation in which more than 90 %, eg. 95 %, 98 % or 99 % of the peptide material in the preparation is that of a polypeptide or polypeptides according to the invention.

The nucleic acid sequences and polypeptides of the invention were originally derived from the barley genome. However, nucleic acid sequences and/or polypeptides of the invention may also be obtained from other eucaryotic genomes, especially other plant genomes. They may be obtained either by conventional cloning techniques or by probing genomic or cDNA libraries with nucleic acid sequences according to the invention. This can be done by any conventional method, such as the methods of Sambrook et al. (Molecular Cloning: A Laboratory Manual; 1989).

The polypeptides of the invention may be linked to a signal sequence capable of directing their secretion from host cells, such as yeast cells, for example cells of S. cerevisiae or S. pombe. Accordingly, nucleic acid sequences, of the invention

may encode such signal sequences in addition to polypeptide sequences having xylose isomerase activity. The nucleic acid sequence encoding the signal sequence must be positioned relative to the sequence encoding the polypeptide having xylose isomerase activity in such a way that the signal sequence is expressed in the host cell and is capable of directing secretion of the polypeptide having xylose isomerase activity. Typically, the nucleic acid encoding the signal sequence will be 5′ to that encoding the polypeptide having xylose isomerase activity. The nucleic acid encoding the signal sequence may be immediately 5′ to that encoding the polypeptide having xylose isomerase activity such that, when the polypeptide is expressed, the signal sequence is immediately N-terminal to the polypeptide having xylose isomerase activity. Alternatively, there may be intervening sequence between the nucleic acid encoding the signal sequence and that encoding the polypeptide having xylose isomerase activity. Further, the nucleic acid encoding the signal sequence must be in the same reading frame as that encoding the polypeptide having xylose isomerase activity. Preferred signal sequences include the Hormoconis resinae glycoamylose signal sequence, though any signal sequence capable of directing secretion of the polypeptide having xylose isomerase activity may be used.

A nucleic acid sequence according to the invention may be included within a vector, suitably a replicable vector, for instance a replicable expression vector.

Such an expression vector comprises an origin of replication so that the vector can be replicated in a host cell such as a bacterial host cell or a yeast host cell. A suitable vector will also typically comprise the following elements, usually in a 5′ to 3′ arrangement: a promoter for the directing expression of the nucleic acid sequence and optionally a regulator of the promoter, a translational start codon, a nucleic acid sequence according to the invention encoding a polypeptide having xylose isomerase activity.

The vector may also contain one or more selectable marker genes, for example an ampicillin resistance gene for the identification of bacterial transformants or a marker gene that allows selection of yeast transformants. Optionally, the vector may also comprise an enhancer for the promoter. The vector may also comprise a polyadenylation signal operably linked 3′ to the nucleic acid encoding the functional protein. The vector may also comprise a transcriptional terminator 3′ to the sequence encoding the polypeptide of the invention.

The vector may also comprise one or more introns or other coding sequences 3′ to the sequence encoding the polypeptide having xylose isomerase activity. The intron or introns may be from barley (the organism from which the sequences of the invention are derived) or the host organism which is to be transformed with the vector or from another eucaryotic organism.

In an expression vector, the nucleic acid sequence of the invention is operably linked to a promoter capable of expressing the sequence. "Operably linked" refers to a juxtaposition wherein the promoter and the nucleic acid sequence encoding the polypeptide having xylose isomerase activity are in a relationship permitting the coding sequence to be expressed under the control of the promoter. Thus, there may be elements such as 5′ non-coding sequence between the promoter and coding sequence. These elements may be native either to barley or to the organism from which the promoter sequence is derived.

Alternatively, the said element or elements may be native to neither the organism from which the promoter sequence is derived nor to barley. Such sequences can be included in the construct if they enhance or do not impair the correct control of the coding sequence by the promoter.

The expression vector may be of any type. The vector may be in linear or circular form. For example, the construct may be

incorporated into a plasmid vector. Those of skill in the art will be able to prepare suitable vectors comprising nucleic acid sequences encoding polypeptides having xylose isomerase activity starting with widely available vectors which will be modified by genetic engineering techniques such as those described by Sambrook et al. (Molecular Cloning: A Laboratory Manual; 1989). So far as plasmid vectors are concerned, two suitable starting vectors are the plasmids PAAH5 (Ammerer (1983): Meth. Enzymol 101, 192-201) and PEVP11 (Hildebrandt et al. (1989): FEBS 243 (2), 137-140), which are widely available.

In an expression vector, any promoter capable of directing expression of a sequence of the invention may be operably linked to the nucleic acid sequence of the invention. Particularly suitable promoters are yeast promoters, for example promoters derived from Kluveromyces spp, Saccharomyces cerevisiae or Schizosaccharomyces pombe. Suitable promoters for the expression of the heterologous genes, such as the barley isomerase gene of the invention, in yeast may be constitutive or regulable. Examples of suitable constitutive promoters are the PDC, PGK, GAPDH, TRP1 and MFα1 promoters. Suitable regulable promoters include the PHO5, ADH1, CUP1, GAL 1, GAL10 and PRB1 promoters. Promoters from viral genes that are expressed in eucaryotic host cells are also suitable. A particularly preferred promoter is the yeast alcohol dehydrogenase (ADH1) promoter.

Typically, nucleic acid sequences according to the invention will be inserted into such vectors in a sense orientation. However, nucleic acid sequences according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA may also be produced by synthetic means. Such antisense RNA may be used in a method of controlling the levels of the protein of SEQ. ID. No. 3 or a protein encoded by a related nucleic acid sequence in a cell.

Vectors according to the invention may be used in vitro, for example for the production of RNA hybridisable to the cDNA. Such vectors may be used to transfect or transform a host cell. Depending on the type of vector, they may be used as cloning vectors to amplify DNA sequences according to the invention or to express this DNA in a host cell.

A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and/or expression of nucleic acid sequences according to the invention, including the DNA SEQ. ID. No. 1 or SEQ. ID. No. 2. The cells will be chosen to be compatible with the vector and may for example be bacterial cells or yeast cells. Transformed or transfected bacterial cells for example E.coli cells, will be particularly useful for amplifying nucleic acid sequences of the invention.

Transformed or transfected yeast cells are particularly preferred for expression of polypeptides according to the invention, which allows them to convert xylose to xylulose. Preferred species of yeast include the distillers' yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. A particularly preferred strain of S. cerevisiae YF135 is strain ALKO246, with the scientific description α leu 2-3, leu 2-112, his 3-11, his 3-15, which strain was deposited under the Budapest Treaty at the Centraalbureau voor Schimmelcultures (CBS) in the Netherlands under deposit number CBS 601.94 on 7th December 1994. A particularly preferred strain of S. pombe is strain ALKO 2185, the scientific description of which is leu 1.32 h-, which strain was deposited under the Budapest Treaty at the Centraalbureau voor Schimmelcultures (CBS) in the Netherlands under deposit number CBS 602.94 on 7 December 1994. The preferred cultivation conditions for the deposited strains are 2 days in a temperature of 30 °C in for example GYPA- or GPYA-medium. Viability tests are performed in the same cultivation condition preferably in GPYA-medium. The strains are preferably stored as freeze dried or in frozen state at -70 °C or lower temperatures.

The cells may be transformed or transfected by any suitable method, such as the methods disclosed by Sambrook et al. (Molecular cloning: A Laboratory Manual; 1989). For example, vectors comprising nucleic acid sequences according to the invention may be packaged into infectious viral particles, such as example retroviral particles. The constructs may also be introduced by electroporation, calcium phosphate precipitation, biolistic methods or by contacting naked nucleic acid vectors with the cells in solution.

In the said nucleic acid vectors with which the host cells are transformed or transfected, the nucleic acid may be DNA or RNA, preferably DNA.

The vectors with which the host cells are transformed or transfected may be of any suitable type. For example the vectors may be able to effect integration of nucleic acid sequences of the invention into the host cell genome or they may remain free in the host cell. Typically, the vectors will be expression vectors, such as a retroviral vector or a DNA expression vector as defined herein. For example, the vector used for transformation construct may be a plasmid vector as defined herein.

The transformed or transfected cells of the invention can be used in a process of production of ethanol production. Cells according to the invention that express polypeptides having xylose isomerase will be capable of converting xylose to xylulose, which is then fermented to ethanol. Such a process of producing ethanol will typically comprise contacting one or more suitable substrates with transformed or transfected cells according to the invention. The thus produced ethanol will typically be recovered by any method known in the art, such as distillation.

Suitable substrates for ethanol production include xylose itself and any compound that the cells can convert into

xylose. For example, suitable, substrates include polymers that contain xylose moieties. Thus, xylan and xyloglucan polymers, which comprise xylose moieties, are suitable substrates. Accordingly, hemicellulosic and lignocellulosic substrates, such as plant biomass, are suitable substrates.

In processes of ethanol production according to the present invention, xylose is typically released from xylan by enzymatic or chemical hydrolysis under acidic or basic conditions, or by heating or by a combination of these techniques. For example xylose can be released from xylan by a combination of acidic or basic hydrolysis and heating, or by heating under pressure. The yeasts typically used in ethanol production are not capable of hydrolysing xylans enzymatically to release xylose although some yeasts are capable of doing so and of metabolising the xylose to ethanol. Such yeasts, when transformed with DNA according to the invention, are included within the scope of the invention.

Although the preferred processes of ethanol production according to the invention comprise contacting transformed or transfected cells with a suitable substrate, other processes of producing ethanol are also possible. For example, a polypeptide according to the invention may be added to a cellular fermentation broth in order to liberate xylose outside the cells, which is then taken up and metabolised by them.

The conditions of the ethanol-producing processes of the invention will typically be adapted and optimised for a suitable level of ethanol production. Such a level of ethanol production will typically be one that is as high as possible without killing or impairing a high proportion of the yeast cells. Ethanol will typically be produced by a fermentation broth that comprises yeast cells according to the invention, water, one or more sources of xylose and other nutrients. In particular, the presence of an appropriate concentration of glucose is desirable as it facilities the growth of the yeast

cells. Also, the presence of Mn^{2+} or Mg^{2+} ions is desirable as one of these ions is necessary to the function of the barley xylose isomerase. Further, the presence of oxygen is desirable, especially in fermentation broths containing S. cerevisiae. The process may be carried at any temperature that facilities ethanol production but temperatures of from 15 to 40 °C are preferred and temperatures of from 30 to 35 °C are particularly preferred.

For the ethanol-producing processes of the invention, it is preferred that transformed or transfected yeast cells, especially those of S. cerevisiae and/or S. pombe are used, including the preferred strains mentioned above.

In the ethanol-producing processes of the invention, the substrate is wholly or partly converted to ethanol, which may be recovered by any suitable means known in the art.

The present invention also provides a process of producing a polypeptide having xylose isomerase activity. Such a process will typically comprise transforming or transfecting host cells with vectors comprising nucleic acid sequences according to the invention and expressing the nucleic acid sequence in these cells. In this case, the nucleic acid sequence will be operably linked to a promoter capable of directing its expression in the host cell. Desirably, such a promoter will be a "strong" promoter capable of achieving high levels of expression in the host cell. It may be desirable to overexpress the polypeptide according to the invention in the host cell. Suitable host cells for this purpose include bacterial cells, for example E. coli cells, and yeast cells, for example those of the preferred species and strains referred to above. The thus produced polypeptide of the invention may be recovered by any suitable method known in the art. Optionally, the thus recovered polypeptide may be purified by any suitable method, for example a method according to Sambrook et al. (Molecular Cloning: A Laboratory Manual).

A further embodiment of the present invention is a process of producing xylulose by contacting a substrate containing xylose with a polypeptide of the invention having xylose isomerase or cells according to the invention that express such a polypeptide. In such a process, the polypeptide according to the invention isomerises xylose to xylulose, which may be recovered by any conventional method.

The nucleic acid sequences of the invention may be used to prepare probes and primers. These will be useful in the isolation of xylose isomerase genes having sequences similar to that of SEQ. ID. No. 2. Such probes and primers may be of any suitable length, desirably from 10 to 100, for example from 10 to 20, 20 to 50 or 50 to 100 bases in length. Two particularly preferred primers are those shown in SEQ. ID. No. 4 and 5.

The present invention also provides antibodies to the polypeptides of the invention, specifically antibodies to the native xylose isomerase protein. These antibodies may be monoclonal or polyclonal. For the purposes of this invention, the term "antibody", unless specified to the contrary, includes fragments of whole antibodies which retain their binding activity for a target antigen. Such fragments include Fv, F(ab') and $F(ab')_2$ fragments, as well as single chain antibodies.

The antibodies may be produced by any method known in the art, such as the methods of Sambrook et al. (Molecular Cloning: A Laboratory Manual; 1989). For example, they may be prepared by conventional hybridoma techniques or, in the case of modified antibodies or fragments, by recombinant DNA technology, for example by the expression in a suitable host vector of a DNA construct encoding the modified antibody or fragment operably linked to a promoter. Suitable host cells include bacterial (for example E. coli), yeast, insect and mammalian cells. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for

example a rat or a rabbit, with a peptide of the invention and recovering immune serum.

The following Examples illustrate the invention.

EXAMPLES

Example 1
Purification, sequencing and characterisation of the xylose isomerase protein.

Materials & Methods

Materials

The starting material for the purification of xylose isomerase enzyme was three to four days old malted barley obtained from Lahden Polttimo (Lahti, Finland). All used column materials were Pharmacia Bioproducts. Sorbitol dehydrogenase and NADH were obtained from Boehringer Mannheim. Xylulose, TPCK (N-tosyl-L-phenylalanine chloromethyl ketone)-treated trypsin, polyvinylpyrrolidone (PVP), XAD-4 (an amberlite non-ionic polymeric adsorbent) and Triton X-100 were Sigma products.

Barley seeds (Hordeum vulgare, cv himalaya) were used as a source for mRNA and genomic DNA isolation. Surface-sterilized seeds were first germinated for 3 to 4 days in the dark at 20 °C on 0.7 % water-agar.

Methods

Xylose Isomerase Assay

Xylose isomerase activity was determined by a two-step procedure where the xylulose formed during the first step was reduced in the second step to xylitol by sorbitol dehydrogenase with a concomitant oxidation of NADH (Callens et al. (1982): Enzyme Microb. Technol. 8, 696-700). The standard

assay mixture was at pH 7.2 and contained 10 mM Tris-HCl, 10 mM $MnCl_2$ and 50 mM D-xylose in a 100 µl reaction volume. After one hour incubation at 35 °C the isomerase reaction was stopped by adding 900 µl ice-cold 0.1 M triethanolamine buffer, pH 7.0. The isomerase reaction product, D-xylulose, was subsequently reduced at pH 7.0 to D-xylitol in a reaction vessel containing 33 µg/ml sorbitol dehydrogenase and 16.8 µM NADH. The amount of NADH oxidised was measured with an AKEA-analyzer, thus providing a measure of the amount of xylulose produced, which is proportional to enzyme activity.

Protein Assay

Protein concentrations were routinely estimated at 280 nm using adsorptivity coefficient of 0.96 $mg^{-1}cm^{-1}$.

Electrophoretic procedures

SDS polyacrylamide gel electrophoresis (PAGE) of protein fractions was carried out as described by Laemmli (1970: Nature 227, 680-685). Western analysis was performed according to the manufacturer's instructions using a "Proto Blot" system supplied by Promega.

Enzyme purification

In order to choose the material for the enzyme purification, one to six days malted barley seeds were analysed by measuring their total protein content and xylose isomerase activity. The results are shown in Fig. 1. Based on these results, three to four days malted barley seeds were chosen as a starting material for the purification. The enzyme has an obligatory requirement for Mn^{2+} or Mg^{2+}, one or both of which were included in all the buffers used. In order to improve the release of enzyme from the seeds, 0.1 % Triton X was used in the extraction buffer and PVP and XAD-4 were included to minimize detrimental reactions between the phenols and proteins.

Protein purification steps were carried out at 4 °C. Chromatography on Superose 12, Mono Q and Phenylsuperose were performed with FPLC system (Pharmacia). Eluates were monitored at 280 nm.

Step 1. Crude protein extract and ammonium sulphate fractionation

500 to 1000 g of three to four days old malted barley was ground slightly in a pH 7.0 buffer containing 10 mM Tris-HCl, 1 mM cysteine, 1 mM $MnCl_2$ (Buffer A). After breaking the seeds 0.1 % (v/v) Triton X-100, 0.1 % (w/v) PVP and 0.1 % XAD-4 (w/v) were added and the mixture was blended continuously for 60 to 75 minutes at +4 °C. The extract was filtered through cheesecloth and centrifuged for 20 minutes (8000 rpm, GSA rotor, Sorvall centrifuge). The supernatant was subjected to ammonium sulphate fractionation. Protein precipitating between 33 % and 60 % saturation with ammonium sulphate was collected by centrifugation as above and dissolved in 100 to 200 ml of pH 7.0 buffer containing 10 mM Pipes, 1 mM cysteine, and 1 mM $MnCl_2$ (Buffer B).

Step 2. Sephadex G-25 gel filtration

In order to remove phenols and other small molecules, the extract was filtered through a Sephadex G-25 column (90 cm x 5 cm) (Strobaek et al. (1976): Carlsberg Res. Commun. 41, 57-72) which was equilibrated with buffer B including 0.1 M NaCl. The flow rate was 10 ml/min and the protein fraction after the void volume was collected.

Step 3. DEAE ion exchange chromatography

A DEAE (Diethylaminoethyl) Sephadex column (20 x 2.9 cm) was equilibrated with the buffer B including 0.1 M NaCl. The protein fraction from G-25 gel filtration was applied into the gel and the column was washed with the same buffer. The

proteins were eluted with a linear gradient of 0.1 M NaCl to 0.25 M NaCl in buffer B (400 ml). The flow rate was 30 ml/h and fractions (4 ml) containing the xylose isomerase activity were pooled and concentrated by filtering through an Amicon P-30 filter. The concentrated material was dissolved in buffer A and applied into the affinity column.

The xylose isomerase was eluted from DEAE ion exchange column with 0.18 M to 0.22 M NaCl, the peak activity eluting with 0.19 to 0.2 M NaCl.

Step 4. Affinity chromatography

The affinity column (2.0 x 10 cm) was prepared by adding xylose sugar into epoxy-activated Sepharose 6B. The addition was performed according to the manufacturer's instructions. The column was further equilibrated with buffer A. After absorbing the pooled and concentrated crude material from the previous step, the column was washed with buffer A and the enzyme was eluted using the same buffer with 1 M NaCl. Fractions containing xylose isomerase activity were concentrated using a Centricon 30 filtering unit according to the manufacturer's instructions. The affinity column required the presence of Mn^{2+} and worked best in buffer A.

Step 5. Gel filtration on Superose 12

A Superose 12 column was equilibrated with a pH 7.0 buffer containing 20 mM Pipes, 1 mM cysteine, 1 mM $MnCl_2$ (buffer C) including 50 mM NaCl at a flow rate of 25 ml/h. Aliquots containing 5 to 10 mg of protein in 200 µl of running buffer were applied into the column and the elution was performed with the same flow rate. The fractions containing xylose isomerase activity (0.58 ml) were pooled and concentrated with Centricon 30 microconcentrators (Amicon). In the Superose 12 gel filtration, the xylose isomerase activity was eluted from the column in the volume 12.18-13.34 ml (fraction 21-23). The recovery varied from 84 % to 89 % in the conditions used.

Step 6. Ion exchange chromatography on Mono Q

A Mono Q column was equilibrated with buffer C at a flow rate of 1 ml/min. The pooled and concentrated fraction (150 µl) from Superose 12 chromatography was applied into the column and washed with 5 ml of buffer C. The activity was eluted with a linear gradient (20 ml) of 0 to 0.3 M NaCl in buffer C. The pooled Mono Q fraction was run on a 10 % SDS PAGE and stained with Coomassie Brilliant Blue. The fractions (250 µl) containing the isomerase activity were pooled and further used for protein sequencing and antibody production.

The thus purified material was about 90 % homogeneous. The results of the purification procedure are given in Table 1.

Table 1. Purification of xylose isomerase

	Volume (Ml)	Total Activity µmol/h	Total Protein mg	Specific Activity µmol/mg	Purific. Factor
Extract	1000	704	17600	0.04	1
Step 1 (60% AmS)	140	910	3556	0.26	6.5
Step 2 (G-25)	320	448	1964	0.23	5.75
Step 3 (DEAE)	81	356	103	3.48	87
Step 4 (Affinity)	7	210	3.85	54.5	1364
Step 5 (Superose 12)	3.5	178	1.45	122.7	3050
Step 6 (MonoQ)	1.25	146	0.1	1462.0	365.62

Peptide Digestions, Purification and Sequencing

The xylose isomerase enzyme isolated as above was further purified for the peptide digest using Bakerbond C-18 reverse phase column in the HPLC system. Protein was bound in 20 % acetonitrile (ACN)/0.1 % trifluoroacetic acid (TFA)-water

solution. The flow rate was 0.6 ml/min and the peak absorbing at 218 nm (16.91 min) was collected and dried in a vacuum using a Speed Vac-system. This sample was further used in peptide digestions with TPCK-treated trypsin. The digest took place in 1 % ammoniumbicarbonate with 2 % (w/w) trypsin for two hours at 37 °C after which more trypsin 2 % (w/w) was added and the reaction was continued overnight.

The peptides were purified using Bakerbond C-18 reverse phase column and the linear gradient of 0-60 % ACN/0.1 % TFA was run at flow rate of 0.60 ml/min. The peaks absorbing at 214 nm were collected manually and applied to the Beckman 890D amino acid sequencer. The following ten peptide sequences were obtained and all were present in the polypeptide sequence (SEQ. ID. NO. 3) of the enzyme deduced from the xylose isomerase cDNA (SEQ. ID. NO. 2) as shown in figures 9A and 9B. Peptides 431, 439 and 443 each had one amino-acid mismatch deduced from the cDNA sequence.

435

I Phe-Gly-Leu-Thr-Gly-Glu-Phe-Lys
(SEQ. ID. No. 9)

II Glu-Gly-Tyr-Gln-Thr-Leu-Leu-Asn-Thr-Asp-Met-Lys
(SEQ.ID. No. 10)

These two peptides were mixed in a single sample but because they were present in widely differing amounts both sequences could be deduced unambiguously.

434

Tyr-Met-His-Gly-Ala-Ala-Thr-Ser-Pro-Glu-Val-Lys
(SEQ. ID. No. 11)

433

(Ser or Trp)-Tyr-Asn-Ala-Glu-Glu-Val-Ile-Leu-(Val)-Gly-Lys
(SEQ. ID. No. 12 and SEQ. ID. No. 13, respectively)

432
Ala-His-Phe-Glu-Phe-Met-Glu-Lys
(SEQ. ID. No. 14)

431
Gly-Thr-Gly-Gly-Val-Pro-Phe-Gly-Ala-Pro-Thr-Lys
(SEQ. ID. No. 15)

443
Xxa-Xxb-Xxc-Xxd-Glu-Leu-Glu-Thr-Ala-Arg (SEQ. ID. No. 16)

Xxa means Met or Ser, Xxb means Tyr or Pro, Xxc means Ala or His and Xxd means Tyr or His.

439
Met-Lys-Asp-Xxx-Leu-Arg
(SEQ. ID. No. 17)

(Xxx stands for "any amino acid")

440
Asn-Asp-Gly-Leu-Ala-Pro-Gly-Gly-Phe (SEQ. ID. No. 18)

444
Ile-Asn-Tyr-Glu-Gly-Pro-Thr-Ser-Lys (SEQ. ID. No. 19)

Further characterisation of the enzyme

Effect of temperature on enzyme activity

Xylose isomerase activity was measured on the temperature range 25-100 °C (Fig. 3). The results from the assays made above 60 °C were corrected for the non-enzymatic isomerization that occurs at such temperatures. Fig. 2 shows the effect of the temperature on the xylose isomerase activity. The temperature optimum is 60 °C and above 80 °C the enzyme lost its activity. The enzyme was quite stable and preserved its

activity after a 5 hour incubation at 60 °C.

Effect of pH on enzyme activity

The effect of pH on the activity of the enzyme was measured in the pH range 4 to 13 (Fig. 3). The following buffers were used 0.05 M acetate/NaOH (pH 4 to 5.6). 0.02 M Bis-Tris (pH 5.5 to 7.0), 0.02M Pipes (pH 6.5 to 7.5). 0.05 M Triethanolamine/NaOH (pH 7.0 to 8.5), 0.010 M Tris-HCl (pH 7.0 to 9.0), 0.05 M glycine-NaOH (pH 8.5 to 10.0), 1 M sodium carbonate/bicarbonate (pH 9.2 to 10.5), 0.025 M disodiumphosphate/NaOH (pH 11 to 12), 0.05 M NaOH/KCl (pH 11 to 13). Above pH 9.5, corrections were made for the non-enzymatic isomerization of D-xylose that occurs under such conditions. The enzyme preserved its activity in the pH range 5.5 to 10.5. Its activity was high across the pH range of 7.0 to 9.0.

Molecular weight

The molecular weight of the enzyme was determined as 100 000 dalton by its elution volume from a calibrated Superose 12 column. SDS/PAGE of the purified xylose isomerase gave a molecular weight estimation of 50 000 suggesting that the native protein is a dimer with two subunits. Molecular weight calculations based on the cDNA sequence gave the estimate 53 620.

Enzyme Specificity

The specificity of the enzyme for xylose, as opposed to glucose and ribose, was determined by measuring its activity when presented with each of the three sugars. No glucose isomerase or ribose isomerase activity was observed, which suggests that the enzyme has a high specificity for xylose.

Enzyme activity was measured in the following manner in each case. A two-step procedure was employed in which the sugar was reduced by sorbitol dehydrogenase with concommitant

oxidation of NADH (Callens et al. (1982): Enzyme Microb Technol. 8, 696-700) The assay mixture was at pH 7.2 and contained 10 mM Tris-HCl, 10 mM $MnCl_2$ and 50 mM of the relevant sugar in a 100 µl reaction volume. After a one hour incubation at 35 °C, the reaction was stopped by adding 900 µl ice-cold 0.1 M triethanolamine buffer at pH 7.0. The reaction product was subsequently reduced at pH 7.0 in a reaction vessel containing 33 µM NADH. The amount of NADH oxidised was measured with an AKEA analyser, in order to give a measure of enzyme activity.

Example 2

Antibody Production

The purified xylose isomerase protein (Mono Q fraction) was used as an antigen and a rabbit was immunized with 20 µg of protein. The immunization was repeated after two and six weeks with the same amount of protein. Blood samples were collected 7 to 10 days after the injection and the final blood sample was collected 7 to 10 days after the last antigen injection.

Example 3

Isolation of cDNA and gene coding for xylose isomerase

DNA manipulations

Restriction endonuclease digests, Southern blot analysis and other techniques were performed according to standard procedures (see, for example, Sambrook et al.: Molecular Cloning: A Laboratory Manual; 1989).

Isolation of Poly(A) RNA

The scutella from two to three days old germinated barley seeds were separated and powdered in liquid nitrogen in a mortar. pH 9.0 buffer containing 0.05 M Tris-HCl, 0.01 M EDTA,

0.1 M NaCl, 2 % (w/v) SDS, and 2 mg/ml proteinase K was added into the powder (10 ml per g of scutellum). The solution was further homogenized with a Polytron homogenizer and incubated for 30 minutes at 40 to 50 °C. After the incubation, the solution was extracted four to five times with the same volume of a solution containing (1:1 phenol and chloroform) : isoamylalcohol (24:1) and finally with chloroform : isoamylalcohol (24:1). Between the extractions the water phase was separated by centrifugation for 15 minutes (4000 rpm, GSA rotor, Sorvall centrifuge). One tenth volume of 4 M NaCl solution and 2 ml of oligo dT-cellulose was added into the protein-free water phase, mixed for 15 minutes and centrifuged in a table top centrifuge at 3000 rpm. The cellulose was washed three times with a pH 7.9 buffer containing 0.01 M Tris-HCl, 0,4 M NaCl, and 0.2 % SDS. After the washings the cellulose was poured into a column equilibrated with a pH 7.5 buffer containing 0.1 M Tris-HCl, and 0.1 M NaCl. The column was eluted with prewarmed 0.1 M Tris-HCl, pH 7.5 buffer and five 3 ml fractions were collected. Poly(A)-RNA was precipitated by adding one tenth volume of 3 M sodium acetate solution and 2 volumes of cold ethanol and keeping the samples at -20 °C overnight.

Preparation and screening of cDNA library

5 µg of the poly(A) RNA as a template cDNA was synthesized using a Promega Kit (Promega, WI, USA) using the components of cDNA synthesis kit supplied by Boehringer Mannheim and following the supplier's instructions. The double-stranded cDNA was treated with T4 polymerase and ligated with EcoRI-adapters and this was further ligated to the vector. Several libraries were prepared using different vectors including gt11, gt 10 and ZAP vectors which were all packed using a "Gigapack Gold II" packing kit supplied by Stratagene. The screening of the libraries were performed using the oligomers synthesised against the peptide sequences as a probe. The positive phage clones picked by oligomers from the gt11 library were further confirmed using antibody screening. After growing for three hours the phages were transferred onto a nitrocellulose filter

and hybridized with the antibody using a "Proto Blot" system supplied by Promega.

Isolation of Full-length cDNA by means of PCR (Polymerase chain reaction)

In order to get a full-length cDNA the RACE protocol designed for PCR (Frohman in "PCR Protocols": eds Innis et al., Academic Press, 1990) was used. The first strand of cDNA was synthesised using the anchor (dT)-oligomer (5'-TTACTCGAGAATTCATCGA$(dT)_{17}$-3' (SEQ. ID. No. 4) as a primer instead of pure oligo(dT). cDNA thus obtained was used further as a template in a PCR reaction and the anchor sequence and the known sequence corresponding the nucleotides 620-644

(5'-TTATGGGGAACTGCACAACTTTC-3') (SEQ. ID. No. 5) within the cDNA sequence were used as primers. The reaction mixture and the amplification conditions were as described by Frohman (1990). The fragment synthesized in the PCR reaction was purified by agarose gel electrophoresis and cloned using a "TA" cloning kit supplied by Invitrogen. Three clones were analyzed by sequencing.

Isolation of total genomic DNA

Total genomic DNA was isolated from the shoots of four to five days old malted barley according to the method described by Dellaporta et al. (1983: Plant Mol. Biol. Rep. 1 (4), 19-21). The isolated total DNA was purified on CsCl gradients with ethidium bromide (10 mg/ml). After removing the ethidium bromide by isoamylalcohol extraction the DNA was dialysed against a 0.010 M Tris-HCl, pH 7.5, 0.001 M EDTA buffer.

Isolation of xylose isomerase gene by PCR

Based on Southern hybridizations with cDNA it was concluded that the isomerase gene was several kilobases in length. Therefore, the gene was isolated in two smaller fragments and

the cDNA sequence was used as a basis for the primer sequences in PCR reactions. Using the primers corresponding to nucleotides 5-30 (sense) and 620-645 (antisense) of the cDNA about two kilobase DNA fragment was obtained which was further cloned using Invitrogen "TA" cloning kit. The genomic area corresponding the 3′-end of the gene was obtained using the primers corresponding to nucleotides 620-645 (sense) and 25 nucleotides preceding the poly(A) tail (antisense) in the cDNA. The fragment obtained was 2.5 kilobases in length and was cloned as above. Three clones for each 5′ or 3′ fragment were sequenced.

DNA sequencing

Positive cDNA clones were sequenced using sequencing kits supplied by USB. Sequencing reactions were analyzed in polyacrylamide gel electrophoresis using an LKB electrophoresis system. PCR clones were sequenced using fluorescently labelled primers and sequencing kits supplied by Applied Biosystems. Reactions were analyzed using Applied Biosystems 373 A automatic sequencer.

Oligonucleotide synthesis

The oligonucleotide sequences corresponding to the peptide sequences were synthesized using Applied Biosystems DNA synthesizer 381A.

Isolation and structure of cDNA clones encoding xylose isomerase

Xylose isomerase cDNA clones from the gt11 library were identified by hybridization with the oligomers based on the peptide sequences 435 I, 435 II, 434, 431 and 432 as probes. The positive clones were further confirmed by the antibody screening.

The sequence of a cDNA clone with an insert of 990 bp is

depicted in SEQ. ID. No. 2. This sequence contained the 5′ end of xylose isomerase cDNA up to the EcoR1 restriction site. In this area all except three (435 I, 443 and 440) peptide sequences were found. The library was screened with the oligomers for these missing sequences but without any positive result. In order to obtain the missing part of the cDNA a RACE PCR was performed. Two of the three PCR clones obtained contained the area from the 5′ primer starting point up to the end including the poly A tail. Clones had an insert of 750 bp and this insert contained also the missing three peptide sequences. The peptide sequence 435 I crosses the end of the previously isolated fragment and the beginning of this new fragment.

Assuming that translation starts with methionine, there is a 75 nucleotide non-translated area at the 5′ end of the full-length cDNA. The non-translated area at the 3′-end of the cDNA is 192 nucleotides and contains three poly(A) addition signal sequences and several stop codons. The full-length cDNA is 1710 nucleotides in size which corresponds with the results obtained with Northern hybridization when the first isolated fragment was used as a probe. The 480 amino acids encoded by the cDNA have a molecular weight of 53 620, which is in good agreement with the values obtained for the protein.

Genomic DNA coding for xylose isomerase

Three PCR clones, each containing 5′ or 3′ ends of the gene were sequenced. A genomic fragment of 4473 bp in length corresponding to the xylose isomerase cDNA sequence was revealed. The genomic sequence was interrupted with 20 intron sequences (Fig. 4). In most cases the exon sequences are quite short compared to the intron sequences. The 5′ non-translated area is for example interrupted with a long intron of 400 nucleotides. In SEQ. ID. No. 1 the whole nucleotide sequence of the xylose isomerase gene is shown. In Figure 5 all 20 exon-intron junctions are represented which shows that every intron starts with GT and ends with AG, thus following

the universal exon-intron rule.

Example 4

Transformation of yeasts by the isomerase cDNA

Materials and methods

Microbial strains and plasmids

Escherichia coli strain XL1-Blue (Bullock (1987): Biotechniques 5, 376-378) Stratagene, La Jolla, CA, USA) was used for propagation of plasmids. The cloning vector was pCRTMII (Invitrogen, San Diego, CA, USA). Plasmids for transformations were constructed by using pAAH5 (Ammerer (1983): Meth Enzymol. 101, 192-201) for Saccharomyces cerevisiae and pEVP11 (Hildebrandt et al (1989): FEBS 243(2), 137-140) for Schizosaccharomyces pombe as basic vectors using standard recombinant DNA techniques. Saccharomyces cerevisiae strain YF135 (deposit number CBS 601.94) and Schizosaccharomyces pombe strain ALKO 2185 (deposit number CBS 602.94 were used as recipients for transformations.

The plasmid pALK252 is constructed from pAAH5 by removing the restriction site XhoI.

Growth media and culture conditions

E. coli strains were grown in L-broth (Sambrook et al. (1989): Molecular Cloning: A Laboratory Manual) supplemented with 50 µg/ml ampicillin when needed. Cultures were grown up at 37 °C overnight.

YPD agar slants (1 % yeast extract, 2 % peptone, both from Difco and 2 % glucose from Merck) were used for storing the Saccharomyces and Schizosaccharomyces strains.

The plates and media for Saccharomyces and Schizosaccharomyces

transformations with leucine selection were 2 % glucose (Merck)/0.67 % yeast nitrogen base without amino acids (from Difco) + amino acids lacking leucine (YNBLeu-).

DNA techniques

DNA manipulations were performed by standard techniques (Sambrook et al. (1989): Molecular Cloning: A Laboratory Manual). The restriction enzymes T4 DNA ligase and Klenow fragment of DNA polymerase I were from Boehringer (Mannheim, Germany) and New England Biolabs, MA, USA). Each enzyme was used according to the supplier's recommendation.

Plasmid DNA from E. coli was isolated by using Qiagen columns (Diagen GmbH, Germany) or the Magic Minipreps DNA Purification System (Promega, Madison, WI, USA) according to the supplier's instructions.

DNA fragments for transformations were isolated from low melting point agarose gels (FMC Bioproducts, Rockland, ME, USA) using beta-agarase from New England Biolabs (Beverly, MA, USA).

Transformation of E. coli strain XL1-Blue was performed by the supplier's suggested method (Stratagene, La Jolla, USA).

For Southern blot analysis the DNA was transferred from agarose gels to nylon membranes by VacuGene TM XL apparatus (Pharmacia, Uppsala, Sweden). The labelling of the probes with digoxigenin and hybridization of the filters were done according to the procedures of Boehringer (Mannheim, Germany).

The PCR reactions were performed by Programmable Thermal Controller PTC-100TM (MJ Research Inc., Watertown, Massachusetts, USA) in 100 µl volumes. The reaction mixture contained 0.1 mM of each dNTP (Cetus), 5 ng of each primer and 50 ng of plasmid template in 1 x buffer supplied by Dynazyme. The protocol used was the following: incubation a 95 °C for 5 min

before adding the F-500L Dynazyme (1 unit, Finnzymes, Espoo, Finland) and 100 µl of paraffin oil; denaturation at 95 °C for 1 min, annealing at 50 °C for 1 minute; extension at 72 °C for 2 min, with a total of 30 cycles each comprising all of these steps. The PCR fragments were purified using the "Magic PCR Preps" DNA Purification System (Promega, Madison, WI, USA).

The oligonucleotides used were sythesized using an Applied Biosystems (Foster City, USA) 381A Synthesizer.

Sequencing was performed directly from the plasmids using an Applied Biosystems Model 373A automatic DNA sequencer.

Transformations of Saccharomyces cerevisiae and Schizosaccharomyces pombe were carried out by the method of Elble (1992: BioTechniques 13(1) 18-20).

Example 5

Construction of Saccharomyces cerevisiae and Schizosaccharomyces pombe strains carrying the cDNA coding for the xylose isomerase gene.

The cDNA encoding the barley xylose isomerase was cloned in two fragments. A 1kb gt11 clone was found to encode the amino terminal part of the barley xylose isomerase. The 1 kb EcoRI-insert was initially transferred into the EcoRI site of pBR322 (pALK705). A unique site for the restriction enzyme StuI was found within the 5′ untranslated region. 1kb StuI-EcoRI fragment was isolated from pALK705 and ligated to SmaI-EcoRI-digested Bluescript SK+ (pALK709) (See Figure 6).

The cDNA clone encoding the carboxy-terminal portion of the barley xylose isomerase was cloned by using specific primers and RT-PCR and a 0.7 kb EcoRI fragment was inserted into EcoRI-digested Bluescript SK+ (pALK706).

To obtain a full length cDNA the carboxyterminal portion of

cDNA from plasmid pAlk706 was isolated and subsequently ligated to the EcoRI site of the plasmid pALK709 linearized with EcoRI to generate plasmid pALK710. The correct orientation of the insert was confirmed with DNA sequencing.

The full length cDNA was transferred into a Saccharomyces cerevisiae expression vector under the control of the alcohol dehydrogenase promoter using PCR.

A forward primer containing a BamHI restriction site and covering the initiator ATG codon and seven additional codons 5′ AAAAGGATCCATG AAG GGC GGG GAG CTC CTG GTC 3′ (oligo 51: SEQ ID. No. 6) and a reverse primer (oligo 50: SEQ ID No. 7) 3′ C TAC GAC AAG GTT AGG CGA GAC ATC CCTAGGAATA 5′ covering the 3′ end of the coding region and containing a BamHI restriction site were synthesized and used in the PCR with pALK710 as the template. The PCR product was digested with BamHI and the ends were filled with dNTPs using the Klenow fragment of the DNA polymerase I. The 1.5 kb fragment was isolated and ligated into HindIII digested and blunt-ended pAAH5 to generate plasmid pALK721 (See Figure 7).

The expression vector pEVP11 containing the Schizosaccharomyces pombe adh1 promoter was linearized with BamHI and ligated with the 1.5 kb BamHI digested PCR product, resulting the expression construct pALK720 (See Figure 7).

After propagation in E. coli, the vectors were transformed into yeast cells using the simplified Li-method of Elble (1992: BioTechniques 13(1), 18-20). The transformants were selected by complementation of the leucine auxotrophy of the host strain.

Example 6

Construction of Saccharomyces cerevisiae and Schizosaccharomyces pombe strains carrying the xylose isomerase gene and the Hormoconis resinae glucoamylase signal sequence.

Construction the plasmids pALK724 and pALK725 (Fig 8).

A new PCR primer with NaeI restriction site (5′ ATTAAGCC GGC GGG GAG CTC CTG GTC 3′) was synthesized and called oligo 54 (SEQ. ID. No. 8).

A PCR reaction was conducted using oligos 54 (SEQ. ID. No. 8) and 50 (SEQ. ID. No. 7) using pALK710 as a template. The PCR product was digested with NaeI and BamHI. A 1.5 kb fragment was isolated and ligated into StuI-BamHI digested pALK730 to generate pALK723. pALK730 is a Bluescript SK+, containing the Hormoconis resinae glucoamylase 5′ untranslated region signal sequence and the prosequence in a 120 bp fragment. This signal sequence is capable of directing secretion of heterologous proteins, such as the barley xylose isomerase from yeast cells. The blunt ended insert can be joined in frame after Lys-Arg codons.

The 1.5 kb insert was separated from pALK723 with restriction enzymes ClaI and BamHI, ends were filled with dNTPs and the blunt- ended fragment was ligated into HindIII digested and blunt-ended pAAH5 to generate the expression construct for S. cerevisisae and to BamHI digested and blunt ended pEVP11 to generate the expression construct for S. pombe.

Example 7

Detection of the xylose isomerase gene in transformed yeasts

5 µg of yeast cells from a YNBLeu- 2 % Glc plate were suspended into 80 µl cell-lysis buffer: 50 mM Tris-HCl pH 8.0/ 2.5 M LiCl/4 % (v/v) Triton X-100. 62.5 mM EDTA and 40 µl of phenol, 40 µl of chloroform and 0.1 g washed glass beads were added. After vortexing the suspension for 1.5 min, the broken cells were centrifuged at 15 000 rpm for 2 min. To the supernatant 80 µl of isopropanol was added. The mixtures were kept 10 min at -20 °C, and centrifuged of 13 000 rpm for 10 min at 4 °C. The precipitates were washed with 70 % (v/v) ethanol, and the

remaining ethanol was removed in Speed-vac evaporator. 20 µl 10 mM tris HCl/0.1mM EDTA buffer was added. 1 µl of the suspensions were used in PCR reactions as described above. The primers were the oligos 50 and 51 described above. 10 µl of the PCR products from S. cerevisiae transformed with pALK 720, from S. pombe transformed with pALK 720, from S. cerevisiae transformed with pALK721 and from S. cerevisiae transformed with pALK724 were run on an agarose gel, and stained with ethidium bromide. The resulting bands showed that the xylose isomerase gene was present in the transformed yeast cells. The standards used in this gel were molecular weight marker III from Boehringer Mannheim. The PCR product identified was about 1.4 kb in length which corresponds to the length expected from the sequencing of the cDNA. The length of the full-length cDNA is 1710 base pairs (see above) and the 1.4 kb fragment corresponds to a 1473 base pair length of cDNA lacking the 3' untranslated region and promoter.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Primalco Ltd.
(B) STREET: Valta-akseli
(C) CITY: Rajamäki
(E) COUNTRY: Finland
(F) POST CODE (ZIP): 05200

(ii) TITLE OF INVENTION: Xylose Isomerase

(iii) NUMBER OF SEQUENCES: 19

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hordeum vulgare cv himalaya

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

TCCGCTTAAC CAACCCCGCA CGGGTCGCGG ACCTGGTTTC CTCCTCCGGC GGCCGCCGCG 60

GTGAGCCCCT CCGCCTCCTT CCTCCCATGC TTCTTCCTCC TCCTCCGCAG ACCCCCTCAG 120

TAGCGGGGAT CCCCCGTTGG CGGCAACCAT CCGTCGGCGC CGTCGTGTCG TCGCTTCGCC 180

GTGACAGAAC CCCATCTCAA GTCTTCTTCG TCGCCAACCC GTCCAGTTTG TGGCCGCGGG 240

CACCATGCCT TAGCTTCATG CCGGAGTACC AGTAGTCATA CAAGCTGTTT ACTTTGCAAA 300

AAATAAATTG GGGATCCGTT CCAATCGAGG GCCCGTTCTT GCTTAGGACA GGCTTGGTAG 360

TTAGGACCCG CGGAATTGGA AAGGTTTCGT TCCGGAGATT TGGATCCTCG TACGGAACAC 420

TTAACTTAAC TTTGTCGATT ATAATTTCTC TTGTTTGAAG TTGCAGAGGC CTTGCGCCAT 480

GAAGGGCGGG GAGCTCCTGG TCCTGCTGCT GGCCTCGTCC CTCTGCCTGT CCGCCGCGGT 540

GAGCGATTGC CAGCTTTGCG GCCTTCTTTG CGTGATTGAG ATGCCCATGG TGGTTTGATC 600

TATCAACTTG ATGCGATCCT GGCTCTGATC TGGCTGCTCC TCGCAGGTTG CCGCGCAGGA 660

AACCTGCCCG GCCGACATCG GCGCCAAGTG CACCGATGCC GCCTCCGATG ATTGGGAGGG 720

CGAGTTCTTC CCCGGCATTG ACAAGATCAA CTATGAGGTG AGGCCATTGA CCTTGGACTG 780

GTGCTAAACA GATTATCTTC ACGGTTCGCT TTCAGTTTAA TAGCTTTTGT CCTTGAACAA 840

TTTCGCCGCA GGGTCCTACC AGCAAGAAGC CGCTTTCTTA CAAGTGGTAT AACGCGGAGG 900

AAGTGATCCT CGGAAAGAAA ATGAAGGTAT GGTTTGCTTT GCAGAACCAT GTCTCGTTGG 960

TGGTAACATC ATTTTAATTT GCTTTTTTAC ATCATGTGCT ACAGTAGGAT TATATATTCG 1020

TTTTTGTGCT ATAGGATTGG TTTCGGTTCA GCGTGGCGTT TTGGCATACG TTCCGGGGTA 1080

CTGGAGGAGA TCCCTTTGGT GCACCTACGA AGAACTGGCC TTGGGAGGAT GGCACCAATT 1140

CCTTGGCCAT GGCTAAGAGA AGAAGTATGA GGATTACTTT ATCTTGTTCT TTTAGATCCA 1200

TGGCGTTATG AAATCGCATA TTTTCTGGAA TCAACAACTT TTATGTTTGG ATTTCTGTTT 1260

CAGTGAAAGC TCACTTCGAG TTCATGGAGA AGCTTGGAGT TGAAAGGTGG TGCTTCCATG 1320

ACAGGGACAT CGCCCCTGAT GGCAAAACAC TCGCGGTATG ATTTTTTCAG AATGATTGCT 1380

GAAGTTGCAA GTTGCAAACC TGAGTTAGGG AGTTGGAGAC TGGTAGTCAC CGATATTAGA 1440

GCAGTTTACT TATGATTATT TACAGGAGTA CAGGACTATG TTTTCAGCAC AAGGGATATG 1500

GTTGCTGATG TTGTACGGCT AATTTTATAC TTGCCATGTA ACTAAAAAGA ATACTCCCTC 1560

CGATCCATAT TAATTGTCGC TGATTTACTA AATCAGCGAC AATTAATATG ACTCGGAGGG 1620

AGTACTTGTT TTGGATCATC TATTCAAGGG CACAATAGAA CCAAACTGAT CACAATAATT 1680

TACAAGCAAT AACAAATATA AATGAACTTT ATTCCAGAAG ATTGCTTCTG TCTAAAATAG 1740

CATCTCTGGG TGTCACTGTA GTATATCTGA AAAATGCGTC GCCAAAATAT ATCTCAAGTG 1800

TGACTGTAAT TTGACTATCT TGGCTCCTAG GCCCTGAAAG GTCAGTATTT CTCCGTCTTA 1860

AGGGTCTTCT GTTCATATTG CAGGAAACAA ATGCTAACTT GGATGAGATA GTTGAGCTGG 1920

CAAAGCAACT CCAGGTTACT GCTACTACCT TAAAATGTTT CAACATTCAG GTTTCCTACA 1980

GTACAACTTT TTTCCGCTCT TTGCTCAAAA ATCTACTCTG AACATCATGT AGAGTGAGAC 2040

CAATATAAAG CCATTATGGG GAACTGCACA ACTTTTCATG CATCCACGTT ACATGCACGG 2100

AGCTGCTACT AGGTGGGTAC TAAATTTCAC TCCATTTGTG AATTTGAGTT ATGACTTGAA 2160

ACTAAGCAAG GCCTTCACTG ATGCTGCAGC CCAGAGGTCA AGGTGTATGC TTATGCTGCT 2220

GCTCAAGTGA AGAAAGCTTT GGAGGTGGGT TGTGTCTGAA TATCATAGGA AGCCTTCCTT 2280

TTGGTTAGTT GTTTAACTTG ATTTGTATCC CCTTCCAGGT TACTCACTAC CTAGGCGGTG 2340

AGAACTACGT ATTCTGGGGT GGAAGAGAGG GTTACCAAAC TCTTCTCAAT ACCGATATGA 2400

AGAGGGAACT TGAACATTTG GTATGGGTTG ACATACTTCT TGACTTTTGT TTGTGTTTCT 2460

ACTTGGTGTA GTTTGTAGCA CAAAAATTCT AGATTGGTAA CTTGTTCTTC TGTTGTGCAG 2520

GCTAACTTTC TTCAAGCTGC TGTTAACCAC AAGAAGAAGA TCGGCTTTAA CGGTAATTGC 2580

TTTTGCAGTG CGATAATTTG ATGTTCCTGG GTTCGAGGCT CCCAAATTTT ACAGTAAATA 2640

GGTATCATTG GTCCGCTCTA TGATAACCTT AAGGATACCT GCGGTGCCTG AATCCCCTTT 2700

TTAATTTCAG GAACATTGTT GATAGAGCCT AAGCCACAAG AACCAACAAA GCATCAGTAC 2760

GTTTGTCGTC TGGAAATATA TATTATGGAA CTGTGAACAG GACTGGAAAG AATTATACTT 2820

ATTTTTCTCT TGCAGGTATG ACTGGGATGT TGCAACTACA TTCTCTTTCC TACAGAAGTT 2880

TGGTCTTACA GGTATTTTTT GAAATGTGGC AAGAAAATCA TTTAGTACAA CCCTTCCTGT 2940

TTTTGTATGT TATGTGTGTA TTTGCAACAA TTCTATGCAA AATGAATCAT GTGAATTAAA 3000

CCATGTTTTC CTGACTATCA ACTTAGAACG TTCTATTTTG AAATAAAATG TATTCTTTTT 3060

TCTAATAATA CGGTGATGCA GGGGAATTCA AGATAAATGT TGAGTGCAAC CATGCTACTC 3120

TCTCTGGACA TAGGTCAGTT TCTTGCTCGC CTAAATCCAA TCATAACGGC TTAAGGTGAT 3180

ACTAGTACCA TAACTGCATC TAAAACTTGT GTCGTTCAGC TGCCATCACG AGCTTGAGAC 3240

TGCACGCATT AATGACATTC TTGGAAACAT TGATGCAAAC ACTGGTGATC CACAGGTTGG 3300

TATGTGTATA GTTCATAAAA TGGCTGTGAC TTGTTAAAAT CTCGGTTACA CGCTTAGTAG 3360

CTTAACATTT CAACTCTTAC GTGTCGGCAG GTTGGGACAC GGATGAGTTC CTTACAGACA 3420

TTTCAGAAGC TACCTTGATT ATGTCAAGTG TAGTTAAGAA TGTGAGTGGA ATTAATCATT 3480

TCTTGTACCT TTTCTGAACC AGAAGTCTAT TGTACTGTAT AATGTTTGTT TACCATTTTT 3540

ATTCAGGTAA CAGTACTCAT AATTTGTTCT GTTTGGTACG ATTGCAGGAT GGACTTGCGC 3600

CTGGTGGCTT CAACTTTTAC GCCAAATTGT ATGTTCCCAA ATAGCATTGG TGTTCCATGA 3660

ATTGAGCTTT GGTTCGACAA ATTTTTTGGC TGACATGGAA TTGTTTTGTG GTAACAGGCG 3720

GAGGGAGAGT ACTGATGTTG AGGACCTGTT TATTGCCCAT ATCTCTGGGA TGGACACCAT 3780

GGCCCGCGGC CGCCGCAATG TTGTCAAGCT GATTGAGGTA ACTGGAGACA TCATTAGTTC 3840

ATCACTCGGT AAAATTTGCA CATGCCTTTA CCTAAATGTA AGGTTTGTTT TCTATGTTAT 3900

TAGGATGGTT CCCTGGACGA GCTTGTACGC AAACGCTACC AGAGCTTTGA CACTGAGATT 3960

GGTGCCATGA TCGAGGTACA CAAATTAACA AGTTCGAATT TATTTTCTCA GAAGGATTGA 4020

ATTGGGATTG TGATACTGGA TTATGGTTTC GAACAGGCTG GGAAGGGCGA CTTTGAAACG 4080

CTAGAAAAGA AGGCCTTGGA GTGGGGCGAG CCAACCGTTC CATCGGGCAA ACAGGTAAAC 4140

GGAACGATAA ACCTATGGCA GCTCGTCTTT CACAGCACAA TGCAAATATA TTTCTGCGCC 4200

GATGATTGTT CCTTACGACC TTTTGTTGCC CGCTTCCCTA TGCAGGAATT GGCTGAGATG 4260

CTGTTCCAAT CCGCTCTGTA GATGGCGGCC CACGGTTCTA GGAATAAAAA AGCAAGAGCG 4320

CGACCTTGGA ACGCCCAGCC GTCCTCGTCA CTACAGGCGA TGTTCTATAG TTAGGCCTCC 4380

ATGCAGTGAA CCCTGTAAAC AAACTGCGTG GAGCTGAAAA TAATGTAACC TTATATCAAA 4440

ATTAAACTCG TTCTTCAACA CGGAATTTGG CTT 4473

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGHTH: 1710 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TCCGCTTAAC CAACCCCGCA CGGGTCGCGG ACCTGGTTTC CTCCTCCGGC GGCCGCCGCG 60

TTGCAGAGGC CTTGCGCCAT GAAGGGCGGG GAGCTCCTGG TCCTGCTGCT GGCCTCGTCC 120

CTCTGCCTGT CCGCCGCGGT TGCCGCGCAG GAAACCTGCC CGGCCGACAT CGGCGCCAAG 180

TGCACCGATG CCGCCTCCGA TGATTGGGAG GGCGAGTTCT TCCCCGGCAT TGACAAGATC 240

AACTATGAGG GTCCTACCAG CAAGAAGCCG CTTTCTTACA AGTGGTATAA CGCGGAGGAA 300

GTGATCCTCG GAAAGAAAAT GAAGGATTGG TTTCGGTTCA GCGTGGCGTT TTGGCATACG 360

TTCCGGGGTA CTGGAGGAGA TCCCTTTGGT GCACCTACGA AGAACTGGCC TTGGGAGGAT 420

GGCACCAATT CCTTGGCCAT GGCTAAGAGA AGAATGAAAG CTCACTTCGA GTTCATGGAG 480

AAGCTTGGAG TTGAAAGGTG GTGCTTCCAT GACAGGGACA TCGCCCCTGA TGGCAAAACA 540

CTCGCGGAAA CAAATGCTAA CTTGGATGAG ATAGTTGAGC TGGCAAAGCA ACTCCAGAGT 600

GAGACCAATA TAAAGCCATT ATGGGGAACT GCACAACTTT TCATGCATCC ACGTTACATG 660

CACGGAGCTG CTACTAGCCC AGAGGTCAAG GTGTATGCTT ATGCTGCTGC TCAAGTGAAG 720

AAAGCTTTGG AGGTTACTCA CTACCTAGGC GGTGAGAACT ACGTATTCTG GGGTGGAAGA 780

GAGGGTTACC AAACTCTTCT CAATACCGAT ATGAAGAGGG AACTTGAACA TTTGGCTAAC 840

TTTCTTCAAG CTGCTGTTAA CCACAAGAAG AAGATCGGCT TTAACGGAAC ATTGTTGATA 900

GAGCCTAAGC CACAAGAACC AACAAAGCAT CAGTATGACT GGGATGTTGC AACTACATTC 960

TCTTTCCTAC AGAAGTTTGG TCTTACAGGG GAATTCAAGA TAAATGTTGA GTGCAACCAT 1020

GCTACTCTCT CTGGACATAG CTGCCATCAC GAGCTTGAGA CTGCACGCAT TAATGACATT 1080

CTTGGAAACA TTGATGCAAA CACTGGTGAT CCACAGGTTG CTTGGGACAC GGATGAGTTC 1140

CTTACAGACA TTTCAGAAGC TACCTTGATT ATGTCAAGTG TAGTTAAGAA TGATGGACTT 1200

GCGCCTGGTG GCTTCAACTT TTACGCCAAA TTGCGGAGGG AGAGTACTGA TGTTGAGGAC 1260

CTGTTTATTG CCCATATCTC TGGGATGGAC ACCATGGCCC GCGGCCGCCG CAATGTTGTC 1320

AAGCTGATTG AGGATGGTTC CCTGGACGAG CTTGTACGCA AACGCTACCA GAGCTTTGAC 1380

ACTGAGATTG GTGCCATGAT CGAGGCTGGG AAGGGCGACT TTGAAACGCT AGAAAAGAAG 1440

GCCTTGGAGT GGGGCGAGCC AACCGTTCCA TCGGGCAAAC AGGAATTGGC TGAGATGCTG 1500

TTCCAATCCG CTCTGTAGAT GGCGGCCCAC GGTTCTAGGA ATAAAAAAGC AAGAGCGCGA 1560

CCTTGGAACG CCCAGCCGTC CTCGTCACTA CAGGCGATGT TCTATAGTTA GGCCTCCATG 1620

CAGTGAACCC TGTAAACAAA CTGCGTGGAG CTGAAAATAA TGTAACCTTA TATCAAAATT 1680

AAACTCGTTC TTCAACACGG AATTTGGCTT 1710

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUNCE CHARACTERISTICS:
(A) LENGTH: 484 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 3:

Met Lys Gly Gly Glu Leu Leu Val Leu Leu Leu Ala Ser Ser Leu Cys
1 5 10 15

Leu Ser Ala Ala Val Ala Ala Gln Glu Thr Cys Pro Ala Asp Ile Gly
20 25 30

Ala Lys Lys Thr Asp Ala Ala Ser Asp Asp Trp Glu Gly Glu Phe Phe
35 40 45

Pro Gly Ile Asp Lys Ile Asn Tyr Glu Gly Pro Thr Ser Lys Lys Pro
50 55 60

Leu Ser Tyr Lys Trp Tyr Asn Ala Glu Glu Val Ile Leu Gly Lys Lys
65 70 75 80

Met Lys Asp Trp Phe Arg Phe Ser Val Ala Phe Trp His Thr Phe Arg
85 90 95

Gly Thr Gly Gly Asp Pro Phe Gly Ala Pro Thr Lys Asn Trp Pro Trp
100 105 110

Glu Asp Gly Thr Asn Ser Leu Ala Met Ala Lys Arg Arg Met Lys Ala
115 120 125

His Phe Glu Phe Met Glu Lys Leu Gly Val Glu Arg Trp Cys Phe His
130 135 140

Asp Arg Asp Ile Ala Pro Asp Gly Lys Thr Leu Ala Glu Thr Asn Ala
145 150 155 160

Asn Leu Asp Glu Ile Val Glu Leu Ala Lys Gln Leu Gln Ser Glu Thr
165 170 175

Asn Ile Lys Pro Leu Trp Gly Thr Ala Gln Leu Phe Met His Pro Arg
180 185 190

Tyr Met His Gly Ala Ala Thr Ser Pro Glu Val Lys Val Tyr Ala Tyr
195 200 205

Ala Ala Ala Gln Val Lys Lys Ala Leu Glu Val Thr His Tyr Leu Gly
210 215 220

Gly Glu Asn Tyr Val Phe Trp Gly Gly Arg Glu Gly Tyr Gln Thr Leu
225 230 235 240

Leu Asn Thr Asp Met Lys Arg Glu Leu Glu His Leu Ala Asn Phe Leu
245 250 255

Gln Ala Ala Val Asn His Lys Lys Lys Ile Gly Phe Asn Gly Thr Leu
260 265 270

Leu Ile Glu Pro Lys Pro Gln Glu Pro Thr Lys His Gln Tyr Asp Trp
275 280 285

Asp Val Ala Thr Thr Phe Ser Phe Leu Gln Lys Phe Gly Leu Thr Gly
290 295 300

Glu Phe Lys Ile Asn Val Glu Cys Asn His Ala Thr Leu Ser Gly His
305 310 315 320

Ser Cys His His Glu Leu Glu Thr Ala Arg Ile Asn Asp Ile Leu Gly
325 330 335 340

Asn Ile Asp Ala Asn Thr Gly Asp Pro Gln Val Gly Trp Asp Thr Asp
345 350 355

```
Glu Phe Leu Thr Asp Ile Ser Glu Ala Thr Leu Ile Met Ser Ser Val
        360                 365                 370

Val Lys Asn Asp Gly Leu Ala Pro Gly Gly Phe Asn Phe Tyr Ala Lys
    375                 380                 385

Leu Arg Arg Glu Ser Thr Asp Val Glu Asp Leu Phe Ile Ala His Ile
390                 395                 400                 405

Ser Gly Met Asp Thr Met Ala Arg Gly Arg Arg Arg Asn Val Lys Leu
                410                 415                 420

Ile Glu Asp Gly Ser Leu Asp Glu Leu Val Arg Lys Arg Tyr Gln Ser
            425                 430                 435

Phe Asp Thr Glu Ile Gly Ala Met Ile Glu Ala Gly Lys Gly Asp Phe
        440                 445                 450

Glu Thr Leu Glu Lys Lys Ala Leu Glu Trp Gly Glu Pro Thr Val Pro
    455                 460                 465

Ser Gly Lys Gln Glu Leu Ala Glu Met Leu Phe Gln Ser Ala Leu
470                 475                 480             484
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUNCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 4:

TTACTCGAGA ATTCATCGA 19

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUNCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 5:

TTATGGGGAA CTGCACAACT TTC 23

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 6:

AAAAGGATCC ATGAAGGGCG GGGAGCTCCT GGTC 34

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 7:

CTACGACAAG GTTAGGCGAG ACATCCCTAG GAATA 35

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION SEQ ID NO: 8:

ATTAAGCCGG CGGGGAGCTC CTGGTC 36

(2) INFORMATION SEQUENCE SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Phe Gly Leu Thr Gly Glu Phe Lys
1 5

(2) INFORMATION SEQUENCE SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Glu Gly Tyr Gln Thr Leu Leu Asn Thr Asp Met Lys
1 5 10

(2) INFORMATION SEQUENCE SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Tyr Met His Gly Ala Ala Thr Ser Pro Glu Val Lys
1 5 10

(2) INFORMATION SEQUENCE SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Ser Tyr Asn Ala Glu Glu Val Ile Leu Val Gly Lys
1 5 10

(2) INFORMATION SEQUENCE SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Trp Tyr Asn Ala Glu Glu Val Ile Leu Val Gly Lys
1 5 10

(2) INFORMATION SEQUENCE SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNASS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Ala His Phe Glu Phe Met Glu Lys
1 5

(2) INFORMATION SEQUENCE SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNeSS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Gly Thr Gly Gly Val Pro Phe Gly Ala Pro Thr Lys
1 5 10

(2) INFORMATION SEQUENCE SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Xxa Xxb Xxc Xxd Glu Leu Glu Thr Ala Arg
1 5 10

(2) INFORMATION SEQUENCE SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Met Lys Asp Xxx Leu Arg
1 5

(2) INFORMATION SEQUENCE SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Asn Asp Gly Leu Ala Pro Gly Gly Phe
1 5

(2) INFORMATION SEQUENCE SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: aminohappo
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Ile Asn Tyr Glu Gly Pro Thr Ser Lys
1 5

CLAIMS

1. An isolated nucleic acid sequence which codes upon expression in a procaryotic or eucaryotic host cell for a polypeptide having xylose isomerase activity, which nucleic acid sequence is selected from:

a) the nucleic acid sequences shown in SEQ. ID. No. 1 and SEQ. ID. No. 2 or the complementary strands thereof;

b) nucleic acid sequences which hybridise to the sequences defined in (a) above;

c) nucleic acid sequences which, but for the degeneracy of the genetic code, would hybridise to the sequences defined in (a) or (b) above and which code for the same polypeptides as those defined in (a) or (b) above.

2. A nucleic acid sequence according to claim 1 that also codes for a signal sequence capable of directing the secretion of the said polypeptide in a host cell, said signal sequence being located 5′ to the nucleic acid sequence defined in (a) (b) or (c) in claim 1 and in the same reading frame as it.

3. A polypeptide encoded by a nucleic acid sequence according to claim 1 or claim 2.

4. A polypeptide according to claim 3 comprising the sequence shown in SEQ. ID. No. 3.

5. A vector comprising a sequence according to claim 1 or claim 2 operably linked to a promoter capable of directing expression of the said sequence in a host cell.

6. A vector according to claim 5 wherein the promoter is capable of directing expression a yeast cell.

7. A vector according to claim 6 wherein the promoter is capable of directing expression in a cell of Saccharomyces cerevisiae or Schizosaccharomyces pombe.

8. Cells transformed or transfected with a vector according to any one of claims 5 to 7.

9. Cells according to claim 8 that are yeast cells.

10. Cells according to claim 9 that are cells of Saccharomyces cerevisiae or Schizosaccharomyces pombe.

11. A process of producing ethanol which comprises:

(i) contacting cells according to any one of claims 8 to 10 with a substrate that comprises one or more carbon sources selected from xylose and polymerised xylose moieties;

(ii) culturing the said cells in conditions under which the isomerisation of D-xylose to D-xylulose occurs and under which the D-xylulose is further catabolized to ethanol; and

(iii) recovering the ethanol.

12. An antibody to a polypeptide of claim 3 or claim 4.

13. A process of producing a polypeptide according to claim 3 or claim 4 comprising expressing a nucleic acid sequence according to claim 1 in a cell according to any one of claims 8 to 10.

14. Ethanol produced by a process according to claim 10.

15. A process of producing xylulose by contacting a substrate containing xylose with a polypeptide according to claim 2 or claim 3 or cells according to any one of claims 7 to 9; and recovering the thus produced xylulose.

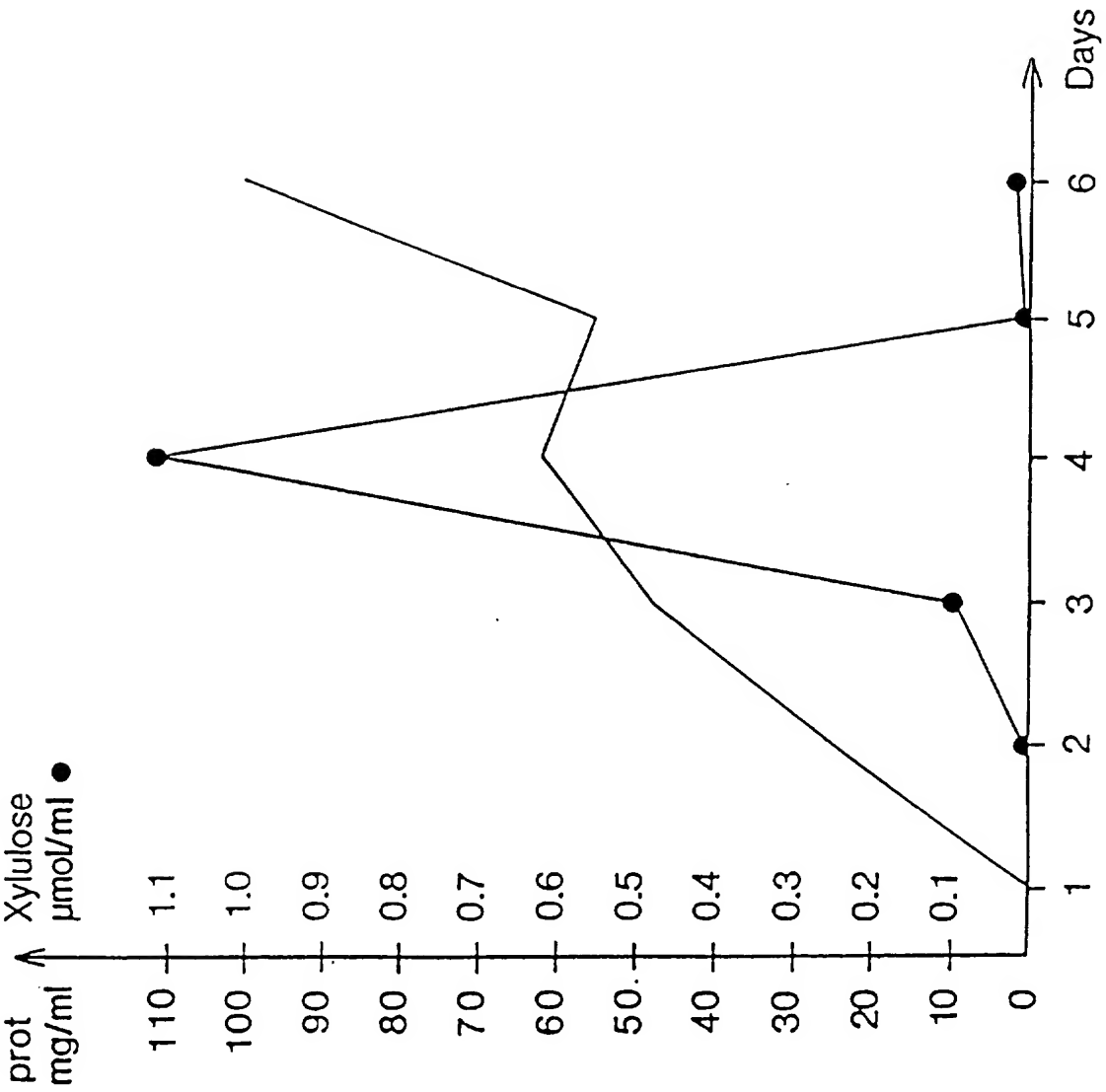

prot mg/ml
Xylulose µmol/ml ●
110
100
90
80
70
60
50
40
30
20
10
0
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1
2
3
4
5
6
Days

FIG. 1

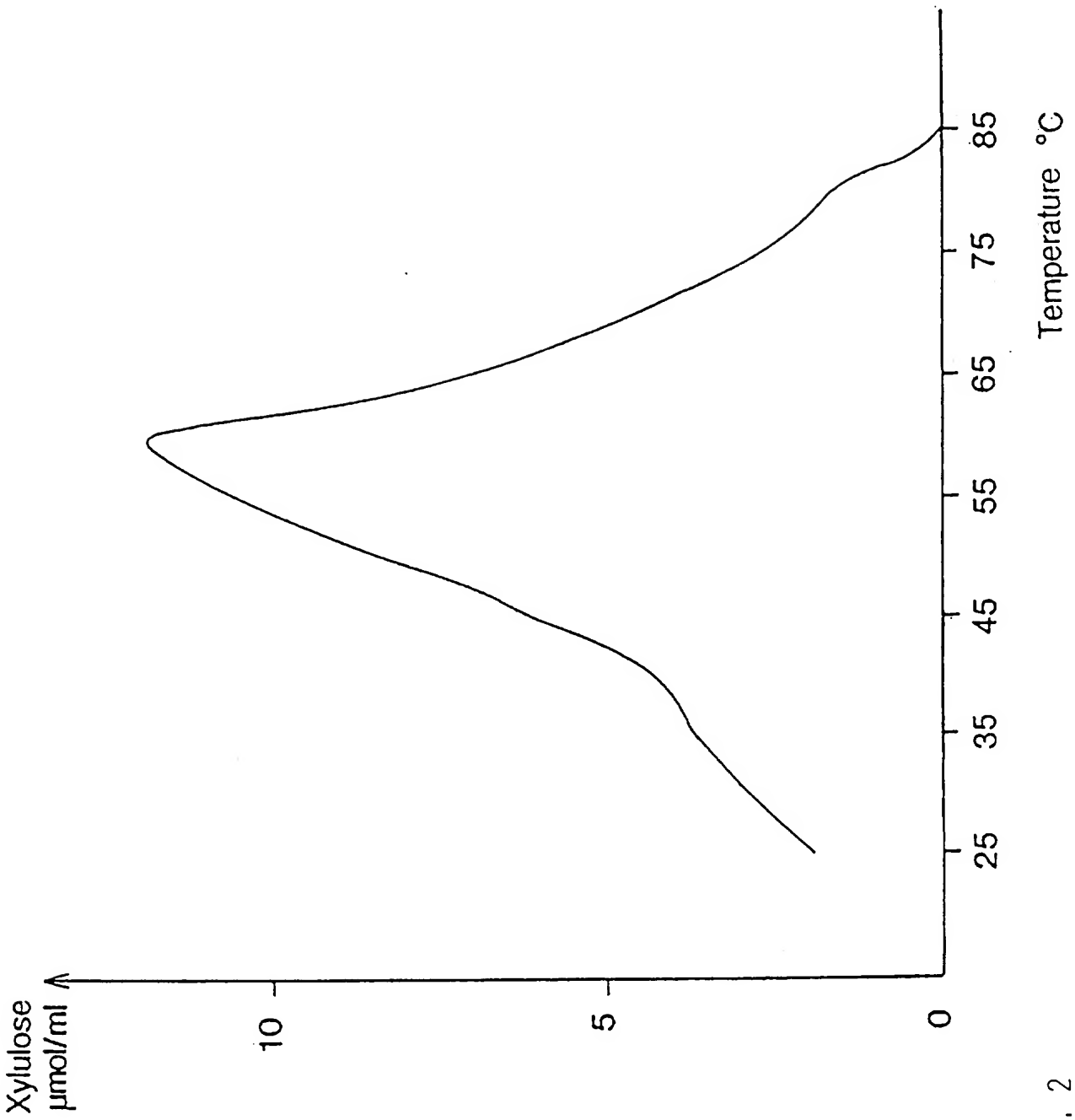


FIG. 2

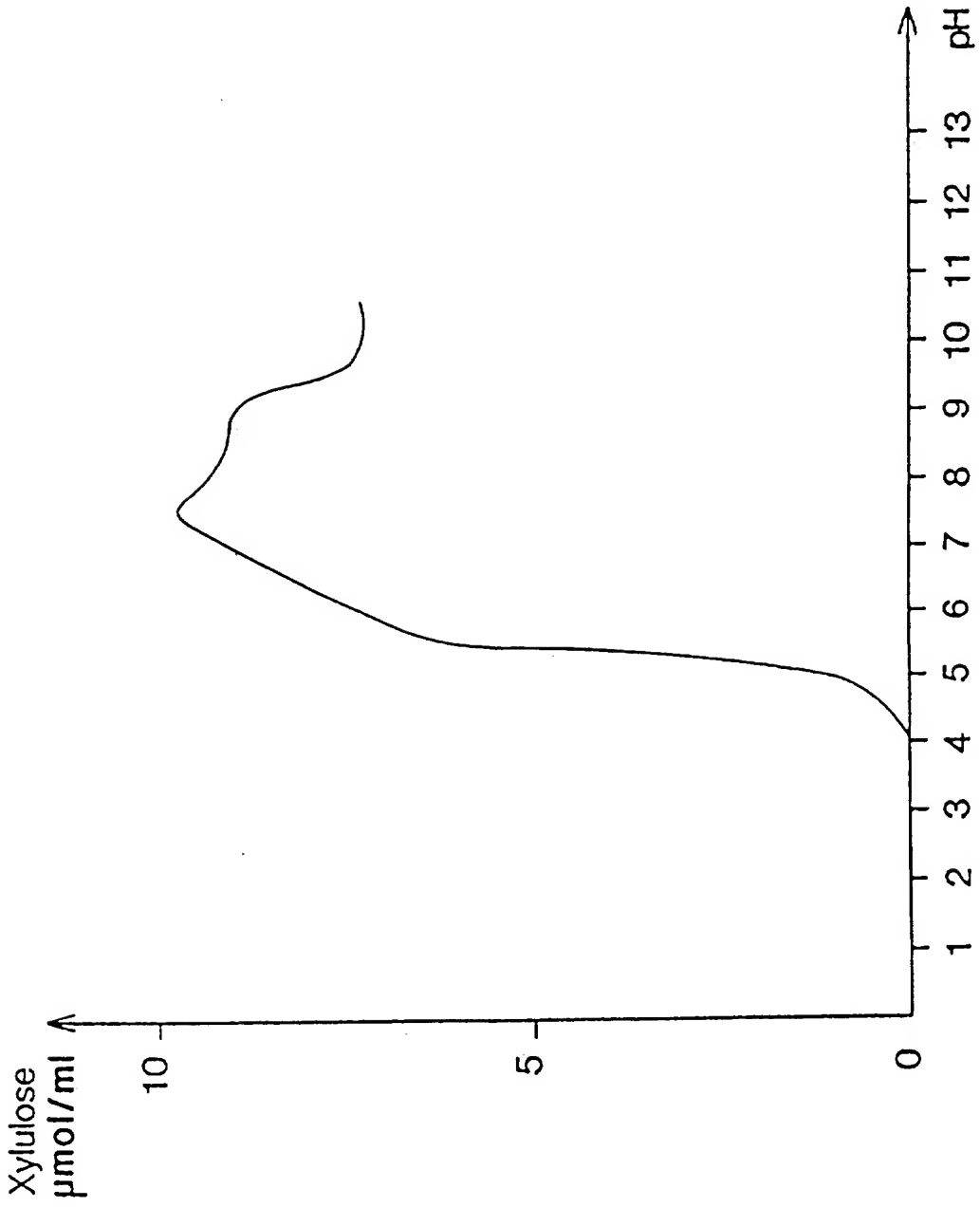
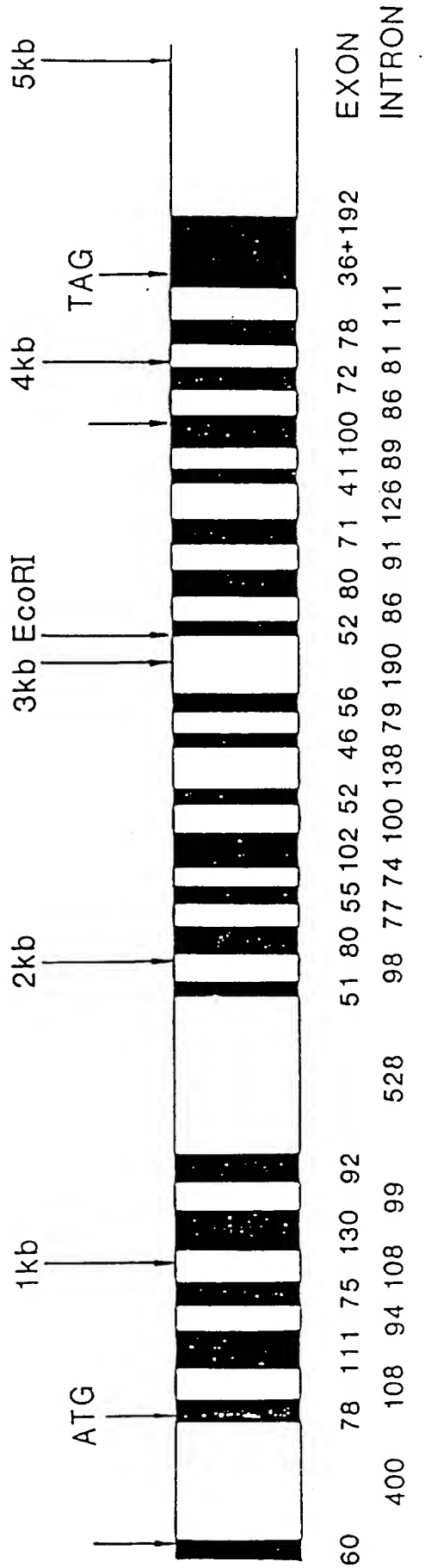


FIG. 3

FIG. 4

XYLOSE-ISOMERASE GENE



EXON /END	INTRON/START	INTRON/END	EXON/START
1. CGCCGCG	GTGAGCCCC	TTGTTTGAAG	TTGCAGAGGC
2. CGCCGCG	GTGAGCGAT	CTCCTCGCAG	GTTGCCGCGC
3. CTATGAG	GTGAGGCCA	TTCGCCGCAG	GGTCCTACCA
4. AATGAAG	GTATGGTTT	TGTGCTATAG	GATTGGTTGC
5. AGAAGAA	GTATGAGGA	TCTGTTTCAG	TGAAAGCTCA
6. ACTCGCG	GTATGATTT	CATATTGCAG	GAAACAAATG
7. ACTCCAG	GTTACTGCT	CATCATGTAG	AGTGAGACCA
8. CTACTAG	GTGGGTACT	GATGCTGCAG	CCCAGAGGTC
9. TTTGGAG	GTGGGTTGT	CCCCTTCCAG	GTTACTCACT
10. ACATTTG	GTATGGGTT	TGTTGTGCAG	GCTAACTTTC
11. TTTAACG	GTAATTGCT	TTAATTTCAG	GAACATTGTT
12. AGCATCA	GTACGTTTG	TCTCTTGCAG	GTATGACTGG
13. CTTACAG	GTATTTTTT	GGTGATGCAG	GGGAATTCAA
14. GACATAG	GTCAGTTTC	TGTCGTTCAG	CTGCCATCAC
15. CAGGTTG	GTATGTGTA	GTGTCGGCAG	GTTGGGACAC
16. TAAGAAT	GTGAGTGGA	ACGATTGCAG	GATGGACTTG
17. CCAAATT	GTATGTTCC	GTGGTAACAG	GCGGAGGGAG
18. GATTGAG	GTAACTGGA	ATGTTATTAG	GATGGTTCCC
19. GATCGAG	GTACACAAA	TTTCGAACAG	GCTGGGAAGG
20. CAAACAG	GTAAACGGA	CCCTATGCAG	GAATTGGCTG

FIG. 5

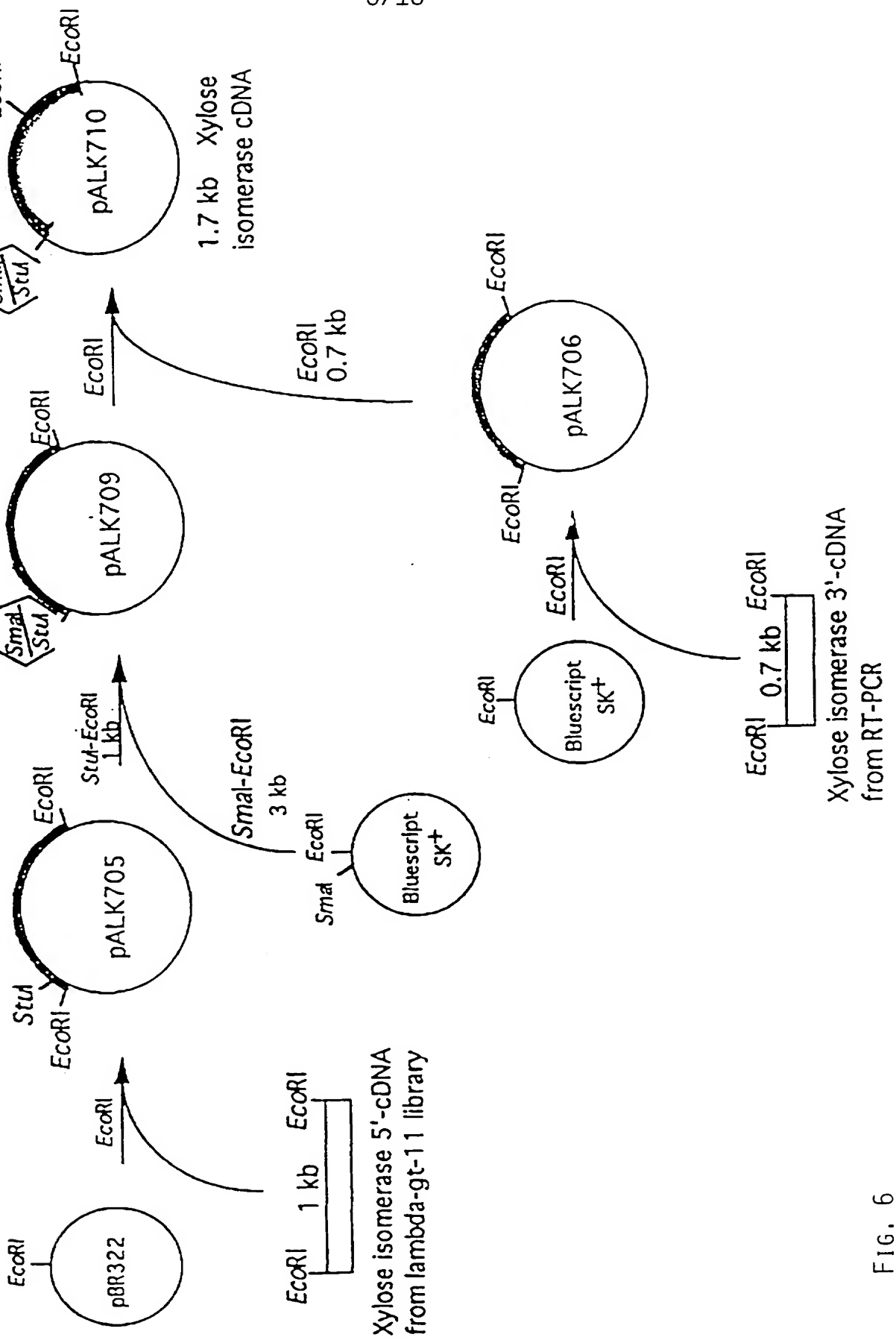

EcoRI
pBR322
EcoRI
1 kb
EcoRI
Xylose isomerase 5'-cDNA from lambda-gt-11 library
EcoRI
Stul
EcoRI
pALK705
Stul-EcoRI 1 kb
SmaI-EcoRI 3 kb
SmaI
EcoRI
Bluescript SK+
SmaI Stul
EcoRI
pALK709
EcoRI
EcoRI 0.7 kb
SmaI Stul
EcoRI
pALK710
EcoRI
1.7 kb Xylose isomerase cDNA
EcoRI
pALK706
EcoRI
EcoRI
Bluescript SK+
EcoRI
0.7 kb
EcoRI
Xylose isomerase 3'-cDNA from RT-PCR

FIG. 6

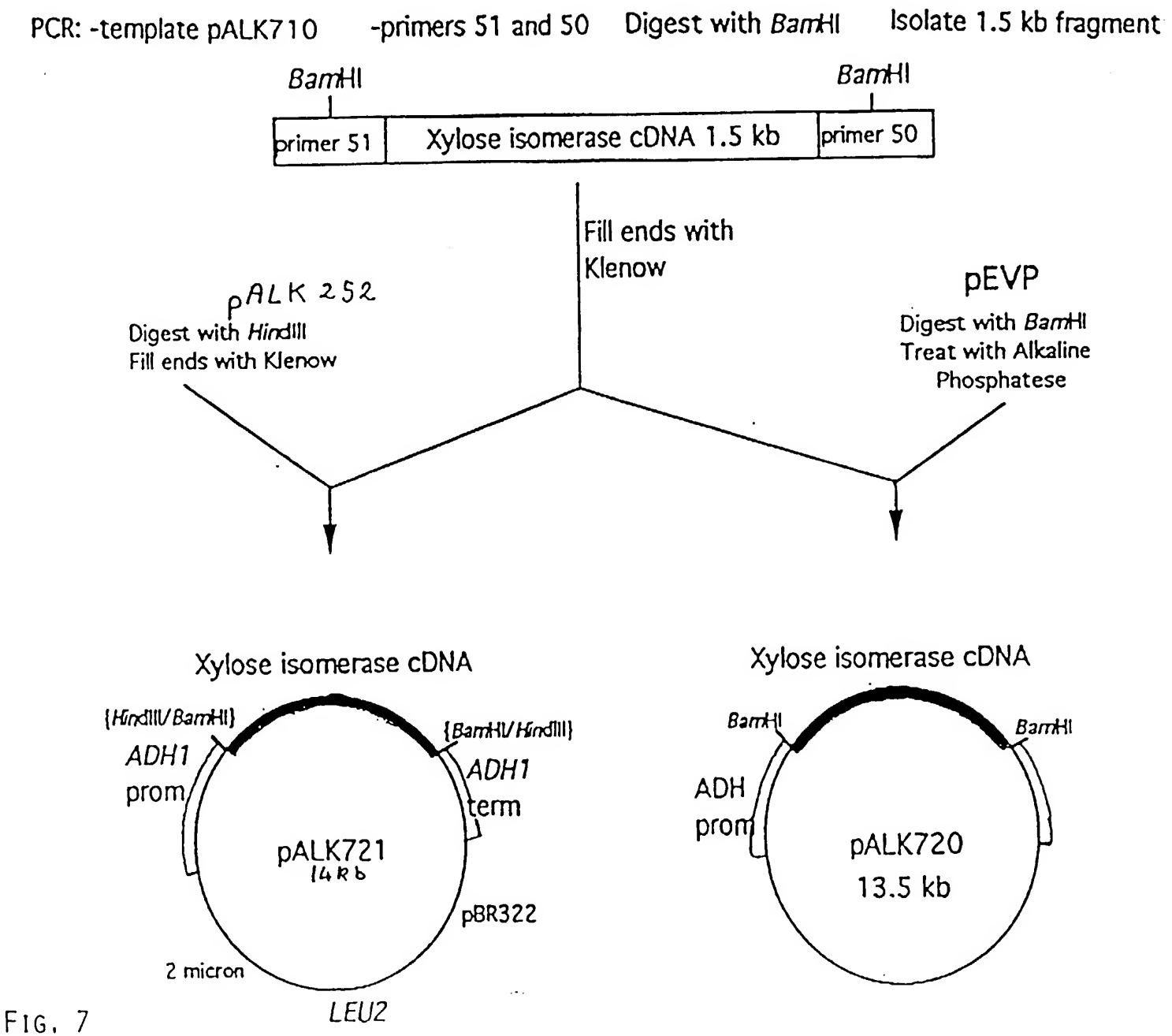

PCR: -template pALK710 -primers 51 and 50 Digest with BamHI Isolate 1.5 kb fragment
BamHI
BamHI
primer 51
Xylose isomerase cDNA 1.5 kb
primer 50
Fill ends with Klenow
pALK 252
Digest with HindIII
Fill ends with Klenow
pEVP
Digest with BamHI
Treat with Alkaline Phosphatese
Xylose isomerase cDNA
{HindIII/BamHI}
{BamHI/HindIII}
ADH1 prom
ADH1 term
pALK721
14 kb
pBR322
2 micron
LEU2
Xylose isomerase cDNA
BamHI
BamHI
ADH prom
pALK720
13.5 kb

FIG. 7

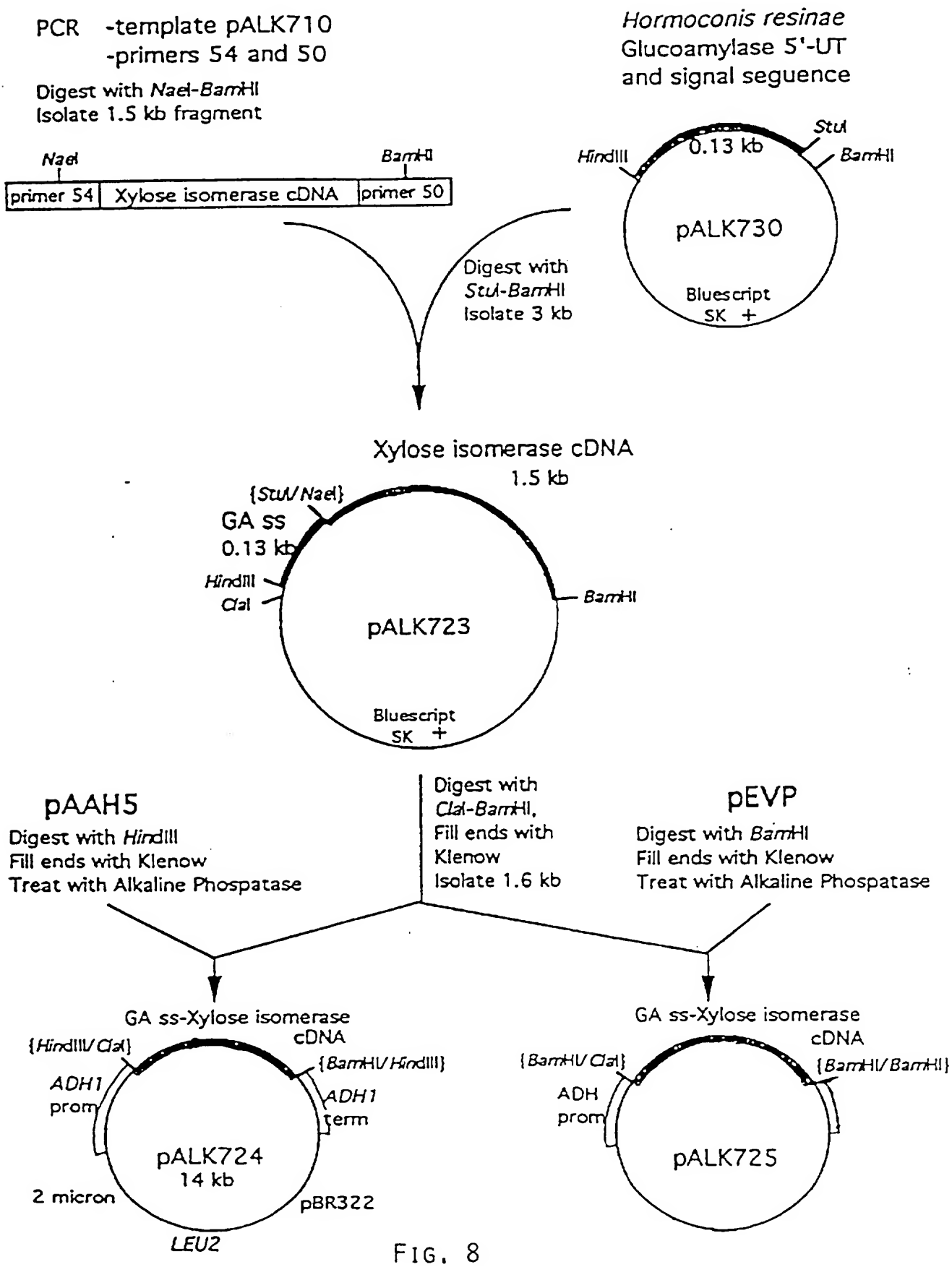

PCR -template pALK710
-primers 54 and 50
Digest with NaeI-BamHI
Isolate 1.5 kb fragment
NaeI
BamHI
primer 54
Xylose isomerase cDNA
primer 50
Hormoconis resinae
Glucoamylase 5'-UT
and signal seguence
StuI
HindIII
0.13 kb
BamHI
pALK730
Bluescript
SK +
Digest with
StuI-BamHI
Isolate 3 kb
Xylose isomerase cDNA
1.5 kb
{StuI/NaeI}
GA ss
0.13 kb
HindIII
ClaI
BamHI
pALK723
Bluescript
SK +
pAAH5
Digest with HindIII
Fill ends with Klenow
Treat with Alkaline Phospatase
Digest with
ClaI-BamHI,
Fill ends with
Klenow
Isolate 1.6 kb
pEVP
Digest with BamHI
Fill ends with Klenow
Treat with Alkaline Phospatase
GA ss-Xylose isomerase
cDNA
{HindIII/ClaI}
{BamHI/HindIII}
ADH1
prom
ADH1
term
pALK724
14 kb
2 micron
pBR322
LEU2
GA ss-Xylose isomerase
cDNA
{BamHI/ClaI}
{BamHI/BamHI}
ADH
prom
pALK725

FIG. 8

```
TCC GCT TAA CCA ACC CCG CAC GGG TCG CGG ACC TGG TTT CCT CCT CCG GCG GCC
GCC GCG TTG CAG AGG CCT TGC GCC ATG AAG GGC GGG GAG CTC CTG GTC CTG CTG
                               Met Lys Gly Gly Glu Leu Leu Val Leu Leu

GTG GCC TCG TCC CTC TGC CTG TCC GCC GCG GTT GCC GCG CAG GAA ACC TGC CCG
Leu Ala Ser Ser Leu Cys Leu Ser Ala Ala Val Ala Ala Gln Glu Thr Cys Pro

GCC GAC ATC GGC GCC AAG TGC ACC GAT GCC GCC TCC GAT GAT TGG GAG GGC GAG
Ala Asp Ile Gly Ala Lys Cys Thr Asp Ala Ala Ser Asp Asp Trp Glu Gly Glu

TTC TTC CCC GGC ATT GAC AAG ATC AAC TAT GAG GGT CCT ACC AGC AAG AAG CCG
Phe Phe Pro Gly Ile Asp Lys Ile Asn Tyr Glu Gly Pro Thr Ser Lys Lys Pro
                                444
CTT TCT TAC AAG TGG TAT AAC GCG GAG GAA GTG ATC CTC GGA AAG AAA ATG AAG
Leu Ser Tyr Lys Trp Tyr Asn Ala Glu Glu Val Ile Leu Gly Lys Lys Met Lys
                433                                                 439
GAT TGG TTT CGG TTC AGC GTG GCG TTT TGG CAT ACG TTC CGG GGT ACT GGA GGA
Asp Trp Phe Arg Phe Ser Val Ala Phe Trp His Thr Phe Arg Gly Thr Gly Gly
                                                    431
GAT CCC TTT GGT GCA CCT ACG AAG AAC TGG CCT TGG GAG GAT GGC ACC AAT TCC
Asp Pro Phe Gly Ala Pro Thr Lys Asn Trp Pro Trp Glu Asp Gly Thr Asn Ser

TTG GCC ATG GCT AAG AGA AGA ATG AAA GCT CAC TTC GAG TTC ATG GAG AAG CTT
Leu Ala Met Ala Lys Arg Arg Met Lys Ala His Phe Glu Phe Met Glu Lys Leu
                                    432
GGA GTT GAA AGG TGG TGC TTC CAT GAC AGG GAC ATC GCC CCT GAT GGC AAA ACA
Gly Val Glu Arg Trp Cys Phe His Asp Arg Asp Ile Ala Pro Asp Gly Lys Thr

CTC GCG GAA ACA AAT GCT AAC TTG GAT GAG ATA GTT GAG CTG GCA AAG CAA CTC
Leu Ala Glu Thr Asn Ala Asn Leu Asp Glu Ile Val Glu Leu Ala Lys Gln Leu

CAG AGT GAG ACC AAT ATA AAG CCA TTA TGG GGA ACT GCA CAA CTT TTC ATG CAT
Gln Ser Glu Thr Asn Ile Lys Pro Leu Trp Gly Thr Ala Gln Leu Phe Met His

CCA CGT TAC ATG CAC GGA GCT GCT ACT AGC CCA GAG GTC AAG GTG TAT GCT TAT
Pro Arg Tyr Met His Gly Ala Ala Thr Ser Pro Glu Val Lys Val Tyr Ala Tyr
      434
GCT GCT GCT CAA GTG AAG AAA GCT TTG GAG GTT ACT CAC TAC CTA GGC GGT GAG
Ala Ala Ala Gln Val Lys Lys Ala Leu Glu Val Thr His Tyr Leu Gly Gly Glu

AAC TAC GTA TTC TGG GGT GGA AGA GAG GGT TAC CAA ACT CTT CTC AAT ACC GAT
Asn Tyr Val Phe Trp Gly Gly Arg Glu Gly Tyr Gln Thr Leu Leu Asn Thr Asp
                                435II
ATG AAG AGG GAA CTT GAA CAT TTG GCT AAC TTT CTT CAA GCT GCT GTT AAC CAC
Met Lys Arg Glu Leu Glu His Leu Ala Asn Phe Leu Gln Ala Ala Val Asn His
```

FIG. 9A

```
AAG AAG AAG ATC GGC TTT AAC GGA ACA TTG TTG ATA GAG CCT AAG CCA CAA GAA
Lys Lys Lys Ile Gly Phe Asn Gly Thr Leu Leu Ile Glu Pro Lys Pro Gln Glu

CCA ACA AAG CAT CAG TAT GAC TGG GAT GTT GCA ACT ACA TTC TCT TTC CTA CAG
Pro Thr Lys His Gln Tyr Asp Trp Asp Val Ala Thr Thr Phe Ser Phe Leu Gln

AAG TTT GGT CTT ACA GGG GAA TTC AAG ATA AAT GTT GAG TGC AAC CAT GCT ACT
Lys Phe Gly Leu Thr Gly Glu Phe Lys Ile Asn Val Glu Cys Asn His Ala Thr
    435I
CTC TCT GGA CAT AGC TGC CAT CAC GAG CTT GAG ACT GCA CGC ATT AAT GAC ATT
Leu Ser Gly His Ser Cys His His Glu Leu Glu Thr Ala Arg Ile Asn Asp Ile

CTT GGA AAC ATT GAT GCA AAC ACT GGT GAT CCA CAG GTT GGT TGG GAC ACG GAT
Leu Gly Asn Ile Asp Ala Asn Thr Gly Asp Pro Gln Val Gly Trp Asp Thr Asp

GAG TTC CTT ACA GAC ATT TCA GAA GCT ACC TTG ATT ATG TCA AGT GTA GTT AAG
GLu Phe Leu Thr Asp Ile Ser Glu Ala Thr Leu Ile Met Ser Ser Val Val Lys

AAT GAT GGA CTT GCG CCT GGT GGC TTC AAC TTT TAC GCC AAA TTG CGG AGG GAG
Asn Asp Gly Leu Ala Pro Gly Gly Phe Asn Phe Tyr Ala Lys Leu Arg Arg Glu
440
AGT ACT GAT GTT GAG GAC CTG TTT ATT GCC CAT ATC TCT GGG ATG GAC ACC ATG
Ser Thr Asp Val Glu Asp Leu Phe Ile Ala His Ile Ser Gly Met Asp Thr Met

GCC CGC GGC CGC CGC AAT GTT GTC AAG CTG ATT GAG GAT GGT TCC CTG GAC GAG
Ala Arg Gly Arg Arg Arg Asn Val Lys Leu Ile Glu Asp Gly Ser Leu Asp Glu

CTT GTA CGC AAA CGC TAC CAG AGC TTT GAC ACT GAG ATT GGT GCC ATG ATC GAG
Leu Val Arg Lys Arg Tyr Gln Ser Phe Asp Thr Glu Ile Gly Ala Met Ile Glu

GCT GGG AAG GGC GAC TTT GAA ACG CTA GAA AAG AAG GCC TTG GAG TGG GGC GAG
Ala Gly Lys Gly Asp Phe Glu Thr Leu Glu Lys Lys Ala Leu Glu Trp Gly Glu

CCA ACC GTT CCA TCG GGC AAA CAG GAA TTG GCT GAG ATG CTG TTC CAA TCC GCT
Pro Thr Val Pro Ser Gly Lys Gln Glu Leu Ala Glu Met Leu Phe Gln Ser Ala

CTG TAG ATG GCG GCC CAC GGT TCT AGG AAT AAA AAA GCA AGA GCG CGA CCT TGG
Leu End

AAC GCC CAG CCG TCC TCG TCA CTA CAG GCG ATG TTC TAT AGT TAG GCC TCC ATG
CAG TGA ACC CTG TAA ACA AAC TGC GTG GAG CTG AAA ATA ATG TAA CCT TAT ATC
AAA ATT AAA CTC GTT CTT CAA CAC GGA ATT TGG CTT
```

FIG. 9B

Applicant's or agent's file reference number	^81PC	International application No.	PCT/FI 96/00070

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 13, line 23 - 28.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet [X]

Name of depositary institution

Centraalbureau voor Schimmelcultures

Address of depositary institution *(including postal code and country)*

Oosterstraat 1
3740 AG Baarn
The Netherlands

Date of deposit	Accession Number
7 December 1994	CBS 601.94

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet [X]

EP The microorganism shall be made available as provided in Rule 28(3) and 28(4) of the implementing regulations of EPC.

FI The microorganism shall be made available as provided in Finnish Patent Law 22 § 7.

Enclosed recognition of receipt and viability statement

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

CA Canada
EP European (all countries)
FI Finland
JP Japan
US USA

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
[X] This sheet was received with the international application	[ ] This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORGANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

Primalco Ltd, Biotec
Valta-akseli
05200 Rajamaki
Finland

name and address of depositor

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT
issued pursuant to Rule 7.1 by the
INTERNATIONAL DEPOSITARY AUTHORITY
identified at the bottom of this page

I. IDENTIFICATION OF THE MICROORGANISM

Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:
ALKO246	CBS 601.94

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under I above was accompanied by:

[X] a scientific description

[ ] a proposed taxonomic designation

(mark with a cross where applicable)

III. RECEIPT AND ACCEPTANCE

This International Depositary accepts the microorganism identified under I above, which was received by it on Wednesday, 7 December 1994 *(date of the original deposit)*[1]

IV. RECEIPT OF REQUEST FOR CONVERSION

The microorganism identified under I above was received by this International Depositary Authority on not applicable *(date of the original deposit)* and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on not applicable *(date of receipt of request for conversion)*

V. INTERNATIONAL DEPOSITARY AUTHORITY

Name: Centraalbureau voor Schimmelcultures

Address: Oosterstraat 1
P.O. Box 273
3740 AG BAARN
The Netherlands

Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):

drs F.M. van Asma
dr M.C. Agterberg

Date: Monday, 15 January 1996

[1] Where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORGANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

Primalco Ltd, Biotec
Valta-akseli
05200 Rajamaki
Finland

name and address of the party to whom the viability statement is issued

VIABILITY STATEMENT
issued pursuant to Rule 10.2 by the
INTERNATIONAL DEPOSITARY AUTHORITY
identified on the following page

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM
Name: Primalco Ltd, Biotec Address: Valta-akseli 05200 Rajamaki Finland	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: CBS 601.94 Date of the deposit or of the transfer: [1] Wednesday, 7 December 1994

III. VIABILITY STATEMENT

The viability of the microorganism identified under II above was tested on Monday, 2 January 1995 [2]. On that date, the said microorganism was

[X] [3] viable

[ ] [3] no longer viable

[1] Indicate the date of the original deposit or, where a new deposit or a transfer has been made, the most recent relevant date (date of the new deposit or date of the transfer).

[2] In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent viability test.

[3] Mark with a cross the applicable box.

IV. CONDITIONS UNDER WHICH THE VIABILITY HAS BEEN PERFORMED [4]	
V. INTERNATIONAL DEPOSITARY AUTHORITY	
Name: Centraalbureau voor Schimmelcultures Address: Oosterstraat 1 P.O. Box 273 3740 AG BAARN The Netherlands	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s): drs F.M. van Asma dr M.C. Agterberg Date: Monday, 15 January 1996

[4] Fill in if the information has been requested and if the results of the test were negative.

Applicant's or agent's file reference number	481PC	International application No	PCT/FI 96/00070

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 13, line 28 - 33.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet [X]

Name of depositary institution

Centraalbureau voor Schimmelcultures

Address of depositary institution *(including postal code and country)*

Oosterstraat 1
3740 AG Baarn
The Netherlands

Date of deposit	Accession Number
7 December 1994	CBS 602.94

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet [X]

EP The microorganism shall be made available as provided in Rule 28(3) and 28(4) of the implementing regulations of EPC.

FI The microorganism shall be made available as provided in Finnish Patent Law 22 § 7.

Enclosed recognition of receipt and viability statement

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

CA Canada
EP European (all countries)
FI Finland
JP Japan
US USA

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
[X] This sheet was received with the international application	[ ] This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORGANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

Primalco Ltd, Biotec
Valta-akseli
05200 Rajamaki
Finland

name and address of depositor

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT
issued pursuant to Rule 7.1 by the
INTERNATIONAL DEPOSITARY AUTHORITY
identified at the bottom of this page

I. IDENTIFICATION OF THE MICROORGANISM

Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:
ALKO2185	CBS 602.94

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under I above was accompanied by:

[X] a scientific description

[ ] a proposed taxonomic designation

(mark with a cross where applicable)

III. RECEIPT AND ACCEPTANCE

This International Depositary accepts the microorganism identified under I above, which was received by it on Wednesday, 7 December 1994 *(date of the original deposit)*[1]

IV. RECEIPT OF REQUEST FOR CONVERSION

The microorganism identified under I above was received by this International Depositary Authority on not applicable *(date of the original deposit)* and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on not applicable *(date of receipt of request for conversion)*

V. INTERNATIONAL DEPOSITARY AUTHORITY

Name: Centraalbureau voor Schimmelcultures

Address: Oosterstraat 1
P.O. Box 273
3740 AG BAARN
The Netherlands

Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):

drs F.M. van Asma
dr M.C. Agterberg

Date: Monday, 15 January 1996

[1] Where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORGANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

Primalco Ltd, Biotec
Valta-akseli
05200 Rajamaki
Finland

name and address of the party to whom the viability statement is issued

VIABILITY STATEMENT
issued pursuant to Rule 10.2 by the
INTERNATIONAL DEPOSITARY AUTHORITY
identified on the following page

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM
Name: Primalco Ltd, Biotec Address: Valta-akseli 05200 Rajamaki Finland	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: CBS 602.94 Date of the deposit or of the transfer: [1] Wednesday, 7 December 1994

III. VIABILITY STATEMENT

The viability of the microorganism identified under II above was tested on Monday, 2 January 1995 [2]. On that date, the said microorganism was

[X] [3] viable

[ ] [3] no longer viable

[1] Indicate the date of the original deposit or, where a new deposit or a transfer has been made, the most recent relevant date (date of the new deposit or date of the transfer).

[2] In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent viability test.

[3] Mark with a cross the applicable box.

IV. CONDITIONS UNDER WHICH THE VIABILITY HAS BEEN PERFORMED [4]	
V. INTERNATIONAL DEPOSITARY AUTHORITY	
Name: Centraalbureau voor Schimmelcultures Address: Oosterstraat 1 P.O. Box 273 3740 AG BAARN The Netherlands	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s): drs F.M. van Asma dr M.C. Agterberg Date: Monday, 15 January 1996

[4] Fill in if the information has been requested and if the results of the test were negative.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI 96/00070

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: C12N 9/92, C12N 15/61
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS, CA, WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Dialog Information Services, File 5, BIOSIS, Dialog accession no. 6657177, Biosis Number: 86123728, Vongsuvanlert V et al: "Purification and characterization of xylose isomerase of a methanol yeast candida-boidinii which is involved in sorbitol production from glucose", Agric Biol Chem 52 (7), 1988, 1817-1824 --	1-15
X	Dialog Information Services, File 5, BIOSIS, Dialog accession no. 11455431, Biosis Number: 98055431, Banerjee S et al: "Xylose metabolism in a thermophilic mould Malbranchea pulchella var. sulfurea TMD-8", Current Microbiology 29 (6), 1994, 349-352 --	1-15

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 May 1996	13 -05- 1996
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Yvonne Siösteen Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI 96/00070

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Dialog Information Services, File 5, BIOSIS, Dialog accession no. 5447087, Biosis Number: 82091890, Zemek J et al: "Xylose isomerase activity of yeasts and yeast-like organisms", Acta Aliment 15 (2), 1986, 111-122	1-15
	--	
A	EP 0483691 A2 (UDAKA, SHIGEZO), 6 May 1992 (06.05.92)	1-15
	--	
A	EP 0194760 A2 (NOVO INDUSTRI A/S), 17 Sept 1986 (17.09.86), page 2, line 20	1-15
	--	

INTERNATIONAL SEARCH REPORT

Information on patent family members

01/04/96

International application No.
PCT/FI 96/00070

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A2- 0483691	06/05/92	JP-A- 5056789	09/03/93
		US-A- 5411886	02/05/95
EP-A2- 0194760	17/09/86	SE-T3- 0194760	
		CA-A- 1316477	20/04/93
		DE-A- 3680802	19/09/91
		JP-A- 61212285	20/09/86
		US-A- 4687742	18/08/87